Technische Universität München
Fakultät für Chemie

Optimal Experimental Designs for the Exploration of Reaction Kinetic Phase Diagrams

Frederic Cassian Victor, Felsen

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Hon.-Prof. Dr. Richard W. Fischer

Prüfer*innen der Dissertation:

1. Prof. Dr. Karsten Reuter
2. Prof. Dr. Klaus Köhler

Die Dissertation wurde am 11.07.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 11.08.2022 angenommen.

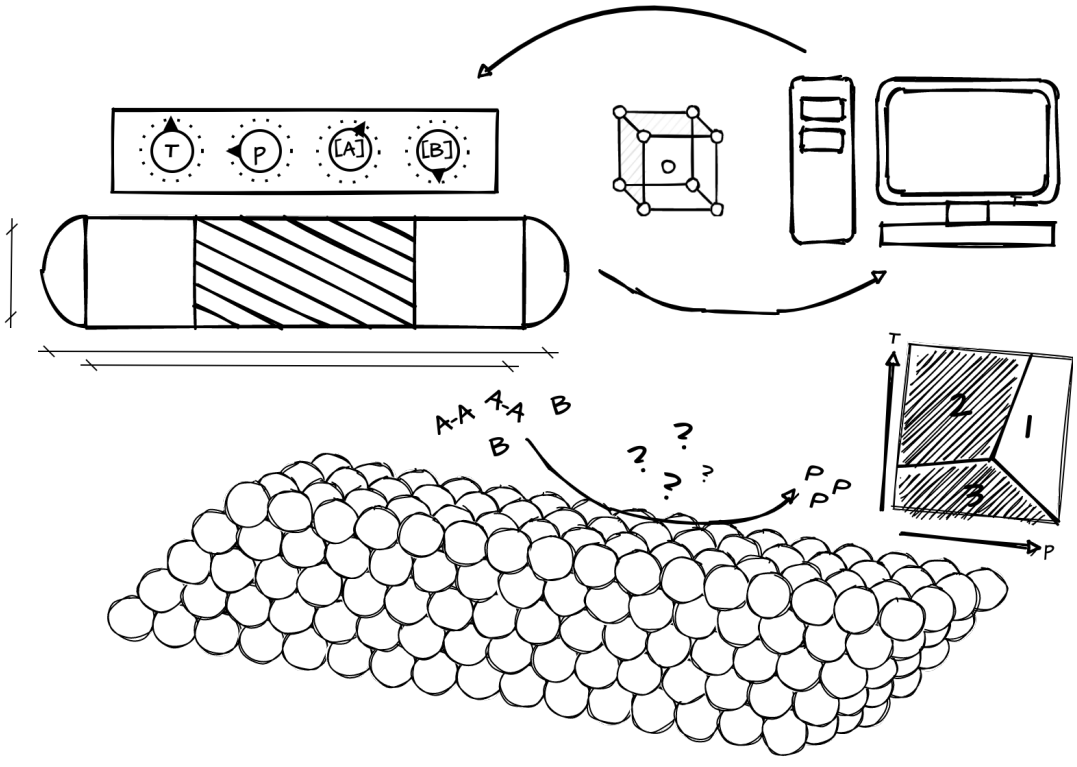
TECHNICAL UNIVERSITY OF MUNICH

DOCTORAL THESIS

Optimal Experimental Designs for the Exploration of Reaction Kinetic Phase Diagrams

Author:
Frederic FELSEN

Supervisor:
Prof. Dr. Karsten REUTER



Preface

The work presented in this doctoral thesis was performed between February 2019 and June 2022 at the Chair of Theoretical Chemistry of the Technical University of Munich (TUM) and the Fritz Haber Institute of the Max Planck Society in Berlin, under the supervision of Professor Karsten Reuter. Parts of this dissertation have been published before in Ref.[1].

Garching, July 2022

Abstract

From the purification of exhaust gases to the large scale industrial production of fertilizers, heterogeneous catalysis is of fundamental importance in various branches of our modern society. A central goal in catalysis research is to understand and control the influence of various experimental parameters on the performance of the reaction. The key to reaching this goal is a systematic way of experimentation both in terms of material synthesis and catalytic performance testing. Statistical design of experiments (DoE) is a mathematical toolbox, which can be used to maximize the information output of a given experimental budget [2]. While traditional DoE is widely applied in industry, academic catalysis research often falls behind. In this thesis, a set of statistical tools especially tailored to tackle problems in chemical kinetics and catalysis is presented by bringing together classical DoE theory with modern statistical learning and optimization methods.

Having a closer look at the catalyst surface, the interplay of various adsorption, diffusion and reaction processes taking place during catalytic reactions can result in a complex kinetic phase diagram [3]. Characteristic for such phase diagrams are transitions with abrupt changes in apparent kinetics. This challenges traditional DoE approaches with their underlying smoothness assumption. The established approaches for modeling the kinetics of catalytic reactions [4, 5] that can cope with such phase transitions suffer from the fact that a detailed understanding of the mechanism is required. Further, these models often also rely on a large number of parameters that are experimentally hardly accessible. Here, making uninformed assumptions can introduce systematic deficiencies into the resulting model. For the investigation of novel catalytic reactions, for which detailed knowledge of the underlying mechanism is not available, a novel data driven regime identification algorithm is proposed.

In order to reduce the bias through *a priori* model assumptions, an algorithm which systematically analyzes the influence of process parameters on the reaction rate to identify effective rate laws without prior knowledge was developed [1]. The proposed method determines relevant model terms from a polynomial ansatz employing well established statistical methods [6]. For the optimization of the model parameters special emphasis is put on the robustness of the results by taking not only the quality of the fit but also the distribution of errors [7] into account in a multi-objective optimization [8]. The flexibility of this approach is demonstrated based on synthetic kinetic data sets from microkinetic models. It could be shown that the kinetics of both the classical HBr reaction and a prototypical catalytic cycle are automatically reproduced based on very limited data sets.

While such rate laws give a reasonable representation of the reaction within one kinetic regime, the low order polynomial approximation will break down approaching phase transitions. Thus, a kinetic regime can also be understood as the range of validity for such an effective regime model. By using local experimental designs in combination with automatically identified kinetic rate laws a local kinetic fingerprint can be created. This way, regions of distinct kinetic behavior are mapped out based on empirically observed data. Combining this local information with unsupervised learning and support vector classification models [9], a global multi-regime kinetic model free of any prior assumptions on the reaction mechanism can be obtained.

Both classical experimental designs as well as space filling sampling techniques [10] are designed for smooth functions of the input parameters. Therefore, the discontinuous behavior at the phase transition requires an alternative sampling approach in order to maximize the efficiency in terms of experimental data. Going beyond established DoE approaches, a modified adaptive experimental design approach [11] is introduced, especially tailored towards modeling experimental regions containing phase transitions. By incorporating the presence of such discontinuities into the model assumption, the position of the real phase transition can iteratively be approached. The potential of this approach is illustrated investigating artificial data sets from a microkinetic model for CO oxidation over RuO_2 [12]. Further, the formulation of this adaptive procedure, which strives for algorithmic equipollence between sets of interrelated continuous and categorical factors opens a wide field of applications going beyond the investigation of chemical kinetics.

Kurzzusammenfassung

Ob bei der Verringerung der Schadstoffemissionen im Straßenverkehr oder bei der großindustriellen Produktion von Düngemitteln, die heterogene Katalyse ist von zentraler Bedeutung für verschiedenste Abläufe in unserer modernen Gesellschaft. Ein wichtiges Ziel in der Katalyseforschung ist es, den Einfluss verschiedener Prozessparameter, wie beispielsweise Druck oder Temperatur, auf den Ablauf der Reaktion zu verstehen und kontrollieren zu können. Voraussetzung dafür ist eine systematische Herangehensweise sowohl in der Synthese der Katalysatormaterialien, als auch bei der Durchführung katalytischer Tests. Ein nützliches Werkzeug in diesem Zusammenhang ist die statistische Versuchsplanung („Design of Experiments“, DoE) [2]. Dabei ist das Ziel, den Informationsgewinn eines begrenzten experimentellen Budgets zu maximieren. In der industriellen Forschung und Entwicklung sind grundlegende Ideen aus der statistischen Versuchsplanung vor allem im Kontext der Prozessoptimierung bereits etabliert. In der akademischen Katalyseforschung wird eine solche, systematische Herangehensweise jedoch in den meisten Fällen vernachlässigt. Diese Arbeit beschreibt eine Reihe von statistischen Methoden, welche speziell auf die Anforderungen der Katalyseforschung und kinetischen Modellierung zugeschnitten sind. Dabei werden klassische Ideen der statistischen Versuchsplanung mit modernen Optimierungsmethoden und Algorithmen kombiniert.

In katalytischen Prozessen resultieren die Wechselwirkungen von verschiedenen Elementarprozessen wie Adsorption, Diffusion oder Reaktion an der aktiven Grenzfläche oftmals in einem komplexen kinetischen Phasendiagramm [3]. Charakteristisch dafür sind vor allem Phasenübergänge welche mit einem abrupten Wechsel des effektiven kinetischen Verhaltens einhergehen. Ein derartig diskontinuierliches Verhalten stellt ein Problem für klassische DoE Ansätze dar, da diese ein glattes Verhalten der Messfunktion voraussetzen.

Komplexe mikrokinetische Modelle sind üblicherweise in der Lage das Verhalten von katalytischen Reaktionen zu beschreiben [4, 5]. Um ein derartiges Modell aufstellen zu können, ist jedoch ein detailliertes Verständnis des Reaktionsmechanismus zwingend notwendig. Die Annahme eines fehlerhaften Mechanismus kann zu systematische Fehlern des resultierenden Modells führen. Aus diesem Grund wird an dieser Stelle ein datenbasierter Algorithmus vorgestellt, welcher speziell für die Analyse neuartiger Reaktionssysteme entwickelt wurde, für welche der zugrundeliegende Mechanismus nicht bekannt ist.

Um systematische Fehler durch *a priori* Annahmen der Modellfunktion zu minimieren, wurde ein Algorithmus entwickelt, welcher systematisch den Einfluss verschiedener Prozessparameter auf die Reaktionsrate untersucht, ohne dabei eine feste funktionelle Form des Modells vorauszusetzen [1]. Dabei werden etablierte statistische Methoden verwendet, um relevante Modellterme aus einem polynomialen Ansatz zu wählen [6]. Die Modellparameter werden im Rahmen einer Pareto-Optimierung [8] bestimmt, wobei neben dem Modellfehler auch die Verteilung der Residuen [7] in Betracht gezogen wird. Anhand synthetischer Datensätze aus mikrokinetischen Modellen für die klassische HBr-Reaktion sowie einen prototypischen katalytischen Zyklus konnte gezeigt werden, dass mit Hilfe eines solchen Algorithmus literaturbekannte Ratengleichungen automatisch reproduziert werden.

Derartige, effektive Ratengesetze sind in der Lage die Kinetik von Reaktionen innerhalb eines Regimes zu beschreiben. Diese polynomiale Näherung der kinetischen Gleichungen bricht jedoch im Bereiche von Phasenübergängen zusammen. Ein kinetisches Regime kann demnach auch als Bereich interpretiert werden, in welchem ein derartiges Ratengesetz eine valide Beschreibung der effektiven Kinetik darstellt. Die Analyse der lokalen Abhängigkeiten der effektiven Kinetik basierend auf lokalen experimentellen Designs kann als Indikator dienen, in welchen Bereichen des experimentellen Bereichs ähnliche Effektivkinetik vorzufinden ist. Mit Hilfe einer Kombination aus dieser lokalen Information mit Methoden des Unsupervised Learning sowie Support Vector basierten Klassifizierungsmodellen [9] kann ein globales Modell der regimeübergreifenden Kinetik ohne jegliche Annahme bezüglich des Reaktionsmechanismus konstruiert werden.

Da sowohl klassische statistische Versuchspläne wie auch raumfüllende Sampling Methoden [10] darauf ausgelegt sind auf glatten Funktionen zu arbeiten, wird für das diskontinuierliche Verhalten nahe der Phasengrenzen ein alternativer Sampling Ansatz benötigt, um die Effizienz des Vorgehens bezügliche experimenteller Daten zu maximieren. Zu diesem Zweck wurde eine modifizierte Variante eines sequentiellen, adaptiven Versuchsplans [11]

entwickelt, welche speziell daran angepasst ist, Diskontinuitäten innerhalb des Design Raumes zu berücksichtigen. Indem das Vorliegen derartiger Phasenübergänge direkt in die Modellannahmen, welche dem adaptiven Algorithmus zugrunde liegen, eingearbeitet wurden, wird die Position der realen Phasengrenze iterativ approximiert. Erste Resultate dieser Methode werden an Hand von mikrokinetisch simulierten Datensätzen für die CO Oxidation über RuO_2 [12] präsentiert. Der Formalismus dieses adaptiven Design Algorithmus, welcher versucht korrelierte kategorische und kontinuierliche experimentelle Faktoren in einem einheitlichen Rahmen zu behandeln, öffnet das Tor zu einem weiten Feld möglicher Anwendung auch jenseits der Untersuchung chemischer Reaktionskinetik.

Contents

List of Abbreviations

DoE	**D**esign **o**f **E**xperiments
FFD	**F**ull **F**actorial **D**esign
GPR	**G**aussian **P**rocess **R**egression
KRR	**K**ernel **R**idge **R**egression
LASSO	**L**east **A**bsolute **S**hrinkage and **S**election **O**perator
LD	**L**ocal **D**esign
LHC	**L**atin **H**yper **C**ube
MAE	**M**aximum **A**bsolute **E**rror
MKM	**M**icro**k**inetic **M**odel
PBD	**P**lacket **B**urman **D**esign
PCA	**P**rincipal **C**omponent **A**nalysis
PPCC	**P**robability **P**lot **C**orrelation **C**oefficient
PSO	**P**article **S**warm **O**ptimizer
QMC	**Q**uasi **M**onte **C**arlo
rbf	**r**adial **b**asis **f**unction
RFR	**R**andom **F**orest **R**egression
RMSE	**R**oot **M**ean **S**quared **E**rror
SAD	**S**equential **A**daptive **D**esign
SPTV	**S**parse **P**olynomial in **T**ransformed **V**ariables
SVC	**S**upport **V**ector **C**lassifier
SVM	**S**upport **V**ector **M**achine
TOF	**T**urn **O**ver **F**requency

Chapter 1

Introduction

"Das wird diesmal total angewandt." - Christoph S.

Understanding how a reaction system evolves over time is a fundamental question in all aspects of chemical research. One of the pioneers in this field was J. H. van't Hoff, who in 1896 laid the groundwork for what is nowadays known as chemical kinetics [13]. In modern research, the study of chemical kinetics is of great importance in various fields of applications. On the one hand the kinetics of proteins and enzymatic reactions in biological systems [14–17] can be crucial in understanding how novel drugs may act on a microscopic scale. On the other hand, knowledge about the kinetics of catalytic reactions [18–20] is relevant in the context of chemical reactor engineering for large scale industrial plants [21]. Looking at a bigger picture, understanding complex reaction networks like the conversion of greenhouse gases could even impact problems like the dependence on fossil fuels and global warming [22].

The modeling of reaction kinetics can be of great value from both an experimental and a theoretical point of view. On the theoretical side, first-principles microkinetic models [23] are a valuable tool in modern catalysis research. These models rely on kinetic and thermodynamic parameters obtained from electronic structure calculations instead of being fit to experimental data. Such models are heavily used in the design and screening of novel catalysts [24–26] as the performance of the candidate materials can directly be assessed *in silico*. Microkinetic models also play a key role in the context of multi-scale modeling [27, 28]. Here, reaction kinetics act as the mesoscopic bridge between microscopic electronic structure calculations and macroscopic computational fluid dynamics. In such a bottom up approach, the goal is to describe macroscopic properties based on fundamental assumptions and theories. In experimental research, the general procedure is reversed. Reaction kinetic models act as a tool to rationalize and compress empirical observations and to elucidate the underlying microscopic mechanism of the process. To this end, complex microkinetic models are of little use, as they require a detailed list of elementary processes as input. Further, the parametrization of microkinetic models requires microscopic quantities like adsorption energies, which are experimentally not easily accessible. Also the huge number of model parameters entering a microkinetic model can be a limitation due to the equally large data set required to estimate them. Further, it needs to be ensured that the available data can actually provide information on all relevant parameters. For example, often times the apparent kinetic behavior of a reaction is only sensitive to the parameters of the rate limiting step [29]. For these reasons, novel chemical systems are generally analyzed using different approaches. Methods for the analysis of experimental data often assume a functional form of a parametric kinetic model and evaluate how well the observed data can be reproduced. Often times, these parametric models are based on careful consideration of classical kinetic theory and therefore, the resulting parameters can be interpreted. If multiple possible kinetic models can be stated, insight about the most probable mechanism can be deduced by comparing the performance of these models [30].

When fitting any model to limited amounts of experimental data, it is crucial to properly construct the set of training points. A frequently encountered shortcoming of experimental data sets is that they are constructed as line scans for the individual process factors like temperature or pressure [31, 32]. A line scan is a one dimensional variation of a single factor

with all other factors kept at a fixed level. While such data sets can provide information about the individual effects of all factors, inference on any kind of interaction is impossible. A better way of constructing these data sets is statistical design of experiments (DoE) [33]. DoE is a toolbox of statistical methods which enables us to probe the effects of various factors on a system in the most efficient way. In the simplest case, standard experimental plans are look up tables which just need to be scaled to the factor levels of the experimental setup. By simultaneously varying multiple factors, such experimental designs provide valuable information about effects and interactions of various experimental parameters.

Coming back to chemical reaction kinetics, the behavior of homogeneous chemical reactions, e.g. in terms of the reaction rate, tends to be smooth as a function of the reaction conditions. In contrast, looking at catalytic reactions at an interface, abrupt changes in apparent behavior can frequently be observed. This phenomenon is called a kinetic phase transition [34]. Phase transitions of this kind can be caused by changes in coverage of the catalyst surface. For example, a poisoning of the catalyst's adsorption sites can result in a dramatic drop in observed reaction rate. By mapping out the position of such transitions as a function of process parameters like temperature or pressure, we obtain a kinetic phase diagram. Microkinetic models are generally capable of capturing this complex behavior, as they take into account the complicated interplay of all elementary reactions [12]. The adequacy of effective analytical models, on the other hand, will break down when crossing a kinetic phase transition, as these models usually are appropriate only for a single regime with approximately smooth behavior. For the investigation of a catalytic reaction, for which multiple effective regimes are expected, there are two options: The first is trying to set up a microkinetic model which can be fitted to experimental data or is based on first principles simulations. Either way, this process will be extremely expensive. Another issue is that assumptions about the reaction mechanism need to be made, which, if not matching the real underlying kinetics, can introduce systematic errors into the model. The alternative option is to first map out the kinetic phase diagram and then fit simple effective models for each of the regimes.

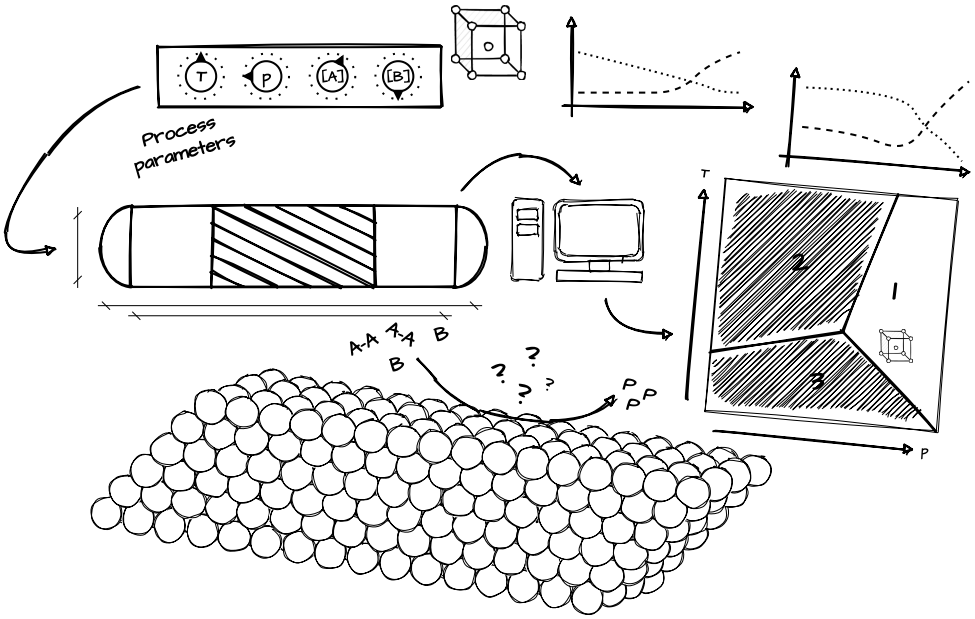


Figure 1.1: Schematic of the identification of kinetic phase diagrams based on empirical kinetic data observed at varying reaction conditions.

Throughout this thesis, the latter option of identifying the kinetic phase diagram first and fitting local effective models second, will be pursued. The main idea is to find a way of locally probing the kinetic system at various reaction conditions and fingerprinting these points according to the local effective kinetics. Based on this information, the critical reaction conditions at which a kinetic phase transition may occur, can be determined. Knowing about the position of the kinetic regimes, we can estimate the range of validity for the regime wide local effective models. Further, by coupling these models, a global multi-regime representation of the system's kinetics can be obtained. In this context, a major question is how to most efficiently sample the space of reaction conditions in order to localize the kinetic phase transitions. To this end, we will employ a sequential adaptive design algorithm and modify it to fit this specific task. A schematic overview of this multi-regime modeling problem is given in Figure 1.1.

Following this introduction, chapter 2 starts with the exhibition of all relevant theoretical approaches and algorithms for the work presented in this thesis. After this theoretical background, the last section of this chapter 2.11 describes the reaction kinetic models used to create artificial data sets throughout this work. Chapter 3 contains the main results of this project, both in terms of developed algorithms and application test cases. The chapter is segmented into three sections, each dealing with one major part of the overall workflow. In section 3.1, a sparse approximation approach to formal reaction kinetics is presented along with its application to classical reaction systems. Section 3.2 deals with the automatized identification of kinetic phase diagrams based on local effective model behavior. Finally, section 3.3 introduces an adaptive design algorithm, which incorporates the procedure for the identification of kinetic regimes in order to improve the data efficiency of this approach. Chapter 4 provides a summary and conclusion of this work followed by an outlook towards future developments and possible applications.

Chapter 2

Theoretical Background

"Das ist ja dann easy." - Christoph S.

2.1 Chemical Reaction Kinetics

The theory of chemical reaction kinetics deals with the temporal evolution of reaction networks. From a mathematical standpoint, this can be described by a stochastic Markov jump process obeying a master equation of the form

$$\frac{\partial P(y,t)}{\partial t} = \int \{W(y|y')P(y',t) - W(y'|y)P(y,t)\}dy'. \tag{2.1}$$

This equation governs the evolution of the probability of the system to reside in state y at time t: $P(y,t)$. This probability is given by the integral over all transitions which lead from or to state y. All processes leading towards y are included in the term $W(y|y')P(y',t)$, which is the product of the probability of being in some other state y' and the transition rate between the two states $W(y|y')$. In the same way, all processes leading from y to some other state are taken into account with a negative sign, indicating that these processes reduce the probability of being in state y [35, 36]. Translating equation 2.1 into chemical terminology, we can think of the state of the system y as being determined by the concentrations and the spatial distribution of all reactant species. Thus, $W(y'|y)$ corresponds to the reaction rates for all elementary steps including y as a reactant or for $W(y|y')$ including y as a product. Under certain assumptions, we can simplify this master equation and arrive at the rate equations from classical kinetics. These assumptions are the following:

(i) The possible states of the system are discrete.

(ii) The temperature must be constant.

(iii) The number of elastic, non-reactive collisions per unit time must be large enough.

(iv) The system must be thermalized.

(v) The reaction mixture must be homogeneous.

Using these assumptions, the rate of a given reaction can generally be expressed in form of rate laws for the time-dependent concentrations $[\mathrm{X_i}](\mathrm{t})$ of the $1 < i < N$ reaction components (reactants and products),

$$\frac{d[\mathrm{X_i}]}{dt} = \sum_j k_j M_{ij} \prod_m ([\mathrm{X}_m]^{s_{jm}}) \quad , \tag{2.2}$$

where k_j are rate constants, M_{ij} are the elements of the stoichiometric matrix and s_{jm} are the stoichiometric coefficients of the reactants [35, 37]. If detailed knowledge on the reaction

mechanism is available, the rate laws for all species involved in the process can be set up according to equation 2.2. Then, by numerically integrating the resulting system of differential equations, the time evolution of the reaction network can be simulated [5]. While rate laws of this form are valid in homogeneous chemistry (in the limit of large system sizes), in heterogeneous systems, we have to introduce another approximation in order to write equation 2.2 in terms of surface coverages. This approximation is the so called mean field approximation (MFA), which assumes statistical independence of the surface adsorption states. Depending on the system at hand, this approximation may introduce severe errors into the kinetic model. Especially for catalytic reactions, in which multiple adsorbed species react with each other, the proximity of these adsorbed species is crucial and assuming a mean coverage will result in large deviations. Models of this kind are generally referred to as microkinetic models (MKMs). In microkinetic modeling, some approaches have been developed to resolve this problem. On the one hand, it is possible to introduce approximate lateral interactions into mean field microkinetic models [38, 39]. The other, more complex approach, is to construct a kinetic Monte-Carlo (kMC) model, which is a general approach to integrate the master equation 2.1. Here, the fact that chemical reactions take place on defined lattice sites at the catalytic surface, simplifies the definition of the individual states of the system. By explicitly operating on a lattice of adsorption sites [4], lateral interactions of the adsorbed species are directly accounted for.

2.2 Linear Regression Modeling

In linear modeling we want to identify the best linear model to explain the variation in an empirical set of points. We assume a linear relation between the independent (or predictor) variables Y and the dependent (or response) variables X plus some error term as both X and Y are random variables[40]:

$$y_{qi} = c_0 + \sum_p c_{pq} x_{pi} + \epsilon_{qi} \tag{2.3}$$

where y_i is the ith observation of the qth response variable, c_0 is a constant offset and the c_{pq} are regression coefficients corresponding to the p predictor variables x_{pi}. Additionally, we have the error term ϵ_{qi}. This describes a relation consisting of two parts, a systematic linear term and a stochastic error term. What we now want to determine are the coefficients c_0 and c_{pq} of the linear part in order to describe the systematic variation in the observed data. In standard linear regression, we make some assumptions on the error in our data. For example, we assume that the predictor variables X are deterministic quantities which do not show stochastic variations upon replication. Further, the random variables ϵ_{qi} are assumed to be independent and equally distributed around an expectation value $E(\epsilon_{qi}) = 0$ [40–42].

2.2.1 Multivariate Linear Regression

The question is now, how to determine the unknown regression coefficients for the linear model. In the case of multiple predictor and response variables, it is useful to rewrite the regression model in a matrix form. If we have N observations, the equation looks as follows:

$$Y_{(Nxq)} = X_{(Nxp)} C_{(pxq)} + \epsilon_{(Nxq)} \tag{2.4}$$

where Y is the response matrix, X is the predictor matrix and C are the unknown coefficients. These coefficients need to fulfill the requirement of minimizing the residual sum of squares of our model:

$$\min_C \left\{ (Y - XC)^T (Y - XC) \right\} \tag{2.5}$$

The resulting, so called, least squares estimator of the regression coefficients $\hat{C}$ is the following:

$$\begin{aligned} \hat{C} &= (X^T X)^{-1} X^T Y \\ \tilde{X} &= (X^T X)^{-1} X^T \end{aligned} \tag{2.6}$$

where $\tilde{X}$ is the Moore-Penrose pseudoinverse [43] of the predictor matrix X. A numerically robust way of obtaining $\tilde{X}$ is the Singular Value Decomposition (SVD) [43]. The SVD can be seen as a generalization of an eigenvalue decomposition for rectangular matrices, which is necessary as we usually have a different number of observations than we have predictor variables.

$$\begin{aligned} X &= U\Sigma V^* \\ \tilde{X} &= V\Sigma^{-1}U^* \end{aligned} \tag{2.7}$$

where U and V are matrices with orthonormal columns and Σ is a diagonal matrix containing the singular values. This way we can directly solve the regression problem by determining $\hat{C} = \tilde{X}Y$.

2.2.2 Regularized Regression

In many cases a simple linear regression will fail to identify an accurate model. This does, however, not necessarily mean that there is no linear relationship in the data. Especially in cases, where we have a high number of predictor variables, we may run into the problem, that the predictor matrix is not of full rank. This is for example the case, if two or more predictors appear to be linearly dependent at a given noise level. Also, it might be that the sensitivity of the response towards certain predictors can not be distinguished from noise and they should therefore be excluded from the final model. In such cases we can apply regularization techniques to obtain better results in our regression. In the following, four common regularization techniques are introduced.

Truncated Singular Value Decomposition (TSVD) As mentioned in section 2.2.1, the SVD can be used to determine the coefficient matrix $\hat{C}$ in linear regression. If the predictor matrix X is rank deficient, several singular values will be zero. If we deal with noisy data, meaning data which is subject to e.g. measurement errors, these singular values will not go to exactly zero, even if two columns of X are perfectly correlated. In a truncated SVD we now only include those singular values, which are larger than some threshold when determining the inverse. For the singular values below this threshold, the corresponding entries in Σ^{-1} (c.f. equation 2.7) are set to zero. This way all singular vectors which correspond to systematic variation are kept, while the noise dimensions are neglected in the back projection to $\tilde{X}$. The threshold values is system specific and needs be determined based on the singular value spectrum of X [43].

Tikhonov Regularization: Ridge Regression Ridge regression is a so called shrinkage method. By introducing a Thikonov regularization term into the least squares cost function of standard linear regression, it penalizes the magnitude of the regression coefficients in the final model:

$$\min_{c_0,c}\left\{\sum_{i=1}^{N}\left(y_i - c_0 - \sum_{j=1}^{p} c_j x_{ij}\right)^2 + \lambda \underbrace{\sum_{j=1}^{p} c_j^2}_{l_2^2}\right\} \tag{2.8}$$

where the penalty term, which is scaled by λ, is applied to all coefficients c_j except for the constant intercept c_0. In otherwise ill posed problems with no unique least squares solution, such a penalty on the l_2^2 norm of the coefficient matrix can lead to a more robust regression by giving preference to a specific solution, namely the one with the smallest euclidian norm [44].

Least Absolute Shrinkage Selection Operator (LASSO) For the purpose of feature selection, ridge regression is not suited, as the ridge solution will always contain all predictor variables. While the penalty on the l_2 will lead to regression coefficients approaching zero, no coefficient will ever be exactly zero. The Least Absolute Shrinkage Selection Operator (LASSO) [6, 45] tries to solve this problem by replacing the l_2^2 norm in the regularization

term by the l_1 norm:

$$\min_{c_0,c}\left\{\sum_{i=1}^{N}\left(y_i - c_0 - \sum_{j=1}^{p} c_j x_{ij}\right)^2 + \lambda \underbrace{\sum_{j=1}^{p}|c_j|}_{l_1}\right\} \tag{2.9}$$

The l_1 norm corresponds to the sum of the absolute values of the coefficients c_j. This has the effect, that some c_j may become exactly zero, if λ is large enough. By setting regression coefficients to zero, the LASSO is able to select only those predictor variables, which contribute significantly to the response. The resuling solution is then referred to as sparse. An alternative way of feature selection would be to iterate over all possible subsets of predictors. This, however, is a combinatorial problem and the computational effort would quickly get out of hand with an increasing number of predictor variables. We can understand the LASSO as computationally feasible approximation to this optimal subset selection [44]. As for ridge regression, the choice of the regularization parameter λ is crucial for the performance of the LASSO. In practice, λ is usually determined via cross validation techniques[46].

Elastic Net Regression The LASSO may run into problems in cases, where multiple predictor variables, which are relevant for the solution, also show a high correlation among each other. To avoid the LASSO picking one of these variables at random, we can apply a different regularization technique, namely the Elastic Net. The Elastic Net is a combination of the LASSO and ridge regression. Both, the l_1 and the l_2^2 norm of the coefficients are added as a penalty term to the target function:

$$\min_{c_0,c}\left\{\sum_{i=1}^{N}\left(y_i - c_0 - \sum_{j=1}^{p} c_j x_{ij}\right)^2 + \lambda_1 \underbrace{\sum_{j=1}^{p}|c_j|}_{l_1} + \lambda_2 \underbrace{\sum_{j=1}^{p} c_j^2}_{l_2^2}\right\} \tag{2.10}$$

Here, we have two regularization parameters λ_1 and λ_2, which control the strength of the LASSO and ridge penalty, respectively. With a correct set of parameters, the combination of the two penalties is able to preserve the variable selection properties of the LASSO while also benefiting from the advantages of ridge regression [47].

2.3 Design of Experiments (DoE)

The field of experimental design deals with the question how to optimally conduct experiments in order to obtain the desired answer with the minimal required effort. DoE is widely applied in various branches of engineering. However, the same fundamental ideas apply to the analysis of any kind of unknown system which is subject to a set of input variables and produces some observable response. These input variables are generally referred to as factors. As indicated in Figure 2.1, the outcome of an experiment can depend on both, factors which can be controlled by the experimenter, and uncontrollable factors, the experimenter can not influence. What we are now interested in, are the effects of the controllable factors at our disposal on the outcome of the experiment, meaning the observable response of the system.

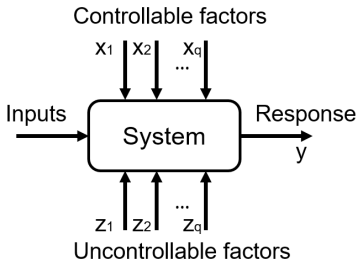


Figure 2.1: Schematic of an unknown system, which is subject to both controllable and uncontrollable factors. Provided a set of inputs it will produce a response, whose values will depend on the setting of the respective influence factors. The illustration is based on Montgomery[48].

The following section will give an overview over the most relevant principles of DoE in the context of this thesis. The discussion is based on relevant literature references, which provide a more extensive and rigorous introduction to the topic [2, 33, 48, 49].

2.3.1 Basic Factorial Experiments

In a general factorial experiment, we are interested in the effects of N controllable factors on some response. In order to gain some inference on these effects, we must conduct multiple experimental runs and vary the settings of all controllable factors whose effect we are interested in. But not only the so called main effects may be significant for the response. Often, interactions of multiple factors should also be taken into account. The question is now, how many of these runs are required and which combinations of factor settings should we be interested in. This problem is basically a linear system of equations. In order for this system to be well determined, the number of observations, i.e. experimental runs, needs to be equal at least to the number of effects we want to estimate. Here, it is important to realize, that the number of effects does grow dramatically with the number of factors, if we also want to account for interactions. Already at the level of pair interactions, the linear increase of the number of effects with the number of factors is accelerated by the binomial coefficient $\binom{n_f}{2}$:

$$n_E = 1 + n_f + \binom{n_f}{2} \tag{2.11}$$

where n_E is the number of effects and n_f the number of factors. This combinatorial growth only gets worse going to higher order effects. If we now want to define an experimental plan, we need to settle on a set of levels for each factor in our experiment. Combinations of these level settings then define an experimental run. The simplest way of setting up such a plan would be to measure all possible combinations of factor settings. This is the so called Full Factorial Design (FFD). For example, the FFD of a three factor experiment with two

levels (-1 and $+1$) would look like follows

$$D = \begin{bmatrix} +1 & +1 & +1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & +1 & -1 \\ -1 & +1 & +1 \\ +1 & -1 & +1 \\ -1 & -1 & -1 \end{bmatrix}, n_E = (n_{levels})^{n_f} = 2^3 = 8 \tag{2.12}$$

The columns of this design matrix D are orthogonal and also each column contains the same number of $+1$ and -1 settings. Thus, the design D is both orthogonal and balanced. These are important requirements in order to be able to independently estimate the effects of individual factors. The FFD always fulfills these requirements. However, the number of experiments for a FFD quickly becomes infeasible, if we increase the number of levels or factors. Therefore, such FFDs are only applicable for simple experimental setups with a low number of factors. The advantage of the FFD is that it contains the maximum amount of information possible about higher order and interaction effects (up to order n_f). For more complex problems we can often reduce the number of experiments by sacrificing information on effects beyond a certain order. Such designs are then generally referred to as fractional factorial designs.

2.3.2 Fractional Factorial Designs

The FFD for an experiment with a large number of factors requires a vast amount of experimental runs. However, higher order interactions make up the bulk of the degrees of freedom in such a system. If, for some reason, we are not interested in such higher order terms, or we can assume these interaction effects to be negligible, we can resort to so called fractional factorial designs to drastically reduce the required number of experiments. By removing several factor settings from the experimental design, also the number of effects, which can be estimated is reduced. Additionally, we no longer can differentiate between the effects of certain factors and interactions terms. This is called a confounding of effects. Looking again at the example of the FFD for a three factor experiment with two levels (this is called the 2^3 factorial design), we expand the columns of the design matrix to see, which factor combinations correspond to higher order effects:

$$\begin{array}{cc|ccc|ccc|c} Run & & A & B & C & AB & AC & BC & ABC \\ 1 & \multirow{4}{*}{$\left[\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right.$} & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ 2 & & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 3 & & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 4 & & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 5 & \multirow{4}{*}{$\left[\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right.$} & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ 6 & & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ 7 & & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ 8 & & -1 & -1 & -1 & +1 & +1 & +1 & -1 \end{array} \tag{2.13}$$

If we now take only the first four lines of this matrix, we can see that e.g. the entries in the A column coincide with those in the BC column. Thus, by only running the first four experiments, we can not differentiate between the effect of A and the interaction BC. The

resolution of a fractional factorial design refers to such a confounding of effects. The most important resolutions for these designs are the following [49]:

Resolution III Designs In these designs no main effects are confounded with each other. They may, however, be confounded with two factor interactions. Further, also two factor interactions can be confounded with each other. Hence, resolution III designs should primarily be used to differentiate between the main effects of various factors. For example, in a screening study, where we want to identify which factors in our experiment are the relevant ones, such designs can be applied.

Resolution IV Designs Using a resolution IV design, we can determine the main effects of all factors more precisely. Now, the main effects are not confounded with any two factor interaction.

Resolution V Designs These designs can be used to differentiate not only between the main effects of all factors, but further also the two factor interactions are not confounded with each other. Confounding with higher order effects is still possible.

Constructing fractional designs, we always have to take the orthogonality and balance of the plan into account. Therefore, it is a common approach to start from a full factorial design for a lower number of factors. For example, replacing the column in the 2^3 factorial design corresponding to a higher order interaction we are not interested in (e.g. ABC) with a new factor D, is a straight forward way of obtaining a fractional plan for four factors. The advantage of fractional factorial plans is, that we always know in advance, which effects we can differentiate and which we can not. Further, by combining multiple fractional factorials we can always extend our plan to achieve a higher resolution [48, 49].

2.3.3 Response Surface Methodology

We can interpret the effects of various factors as coefficients in a linear model, which describes the change of the observed response as a function of the change in a given factor level. In general, we can approximate such a function as a polynomial of the input factors [2]. A polynomial is a linear combination of products and powers of the input variables. Despite containing such nonlinear terms, a polynomial model is still linear in the coefficients, which we identify with the effects of all factors and their respective powers and interactions. Let us consider a simple system with two factors x_1 and x_2 and a single response y. The corresponding polynomial up to second order would look as follows:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2 \tag{2.14}$$

The coefficients of these various polynomial terms can be determined using linear regression. So, if we have performed a set of experiments given an experimental design, we can directly analyze the effects of interest by fitting such a regression model. As previously mentioned, the order up to which the polynomial is expanded, is limited by the amount of data provided. In Response Surface Methodology (RSM) we now want to exploit these models for the response (in higher dimension then referred to as response surfaces) in order to not only describe, but also tune the factors at our disposal. Fitting a second order polynomial to our data enables us to optimize the response using minimization algorithms and hence find the optimal factor settings for our system. Such techniques are commonly employed in process engineering [50, 51]. For example, one could be interested in the optimal feed composition in order to maximize the yield of a chemical reaction. In order to fit second order models, special experimental designs have been developed to estimate second order effects in a efficient way. Common examples for such response surface designs are the Central-Composite Design (CCD) and the Box-Behnken Design (BBD) [2, 48, 52].

2.3.4 Sequential Adaptive Designs

The aforementioned classical experimental designs usually provide a complete set of experiments, which, if performed, allow for an analysis of effects with a certain resolution or provide enough data to estimate the parameters of a given model. These designs often are extensible with additional points. For example, a fold over can be used to transform any resolution III design into a resolution IV design. In contrast, sequential designs start from some initial set of data and suggest subsequent experiments to be performed, one at a time. In this process, all the available data up to this point is taken into account in order to identify the most promising next sample point [11].

The D-Optimality Criterion In order to find the next point in a sequential design, we need to evaluate possible candidates and define a criterion to select the best one. There are different criteria discussed in literature[53], which all relate to the information content of the design in terms of its information matrix. Let us first assume that the model η is scalar and linear in the parameters $\boldsymbol{\theta}$. This means that

$$\begin{aligned} \eta(\boldsymbol{x}, \boldsymbol{\theta}) &= \boldsymbol{\theta}^T \boldsymbol{f}(\boldsymbol{x}) \\ \boldsymbol{\theta}^T &= [\theta_0, \theta_1, ..., \theta_k] \\ \boldsymbol{f}(\boldsymbol{x}) &= [f_0(\boldsymbol{x}), f_1(\boldsymbol{x}), ..., f_k(\boldsymbol{x})]^T \end{aligned} \tag{2.15}$$

where the $f_k(\boldsymbol{x})$ are known functions independent of $\boldsymbol{\theta}$. In this linear parametrization case, the information matrix M can be written as

$$M = \sum_i^N w_i \boldsymbol{f}(\boldsymbol{x}_i) \boldsymbol{f}^T(\boldsymbol{x}_i) \quad with \quad w_i = \frac{1}{b_i^2} \tag{2.16}$$

where b_i^2 is the variance of the measurement at point $\boldsymbol{x}_i$. This information matrix is the inverse of the dispersion matrix of the parameter estimator $\hat{\boldsymbol{\theta}}$:

$$D(\hat{\boldsymbol{\theta}}) = M(\hat{\boldsymbol{\theta}})^{-1}. \tag{2.17}$$

It can be shown that the best linear estimate of the parameters, $\hat{\boldsymbol{\theta}}$, minimizes the determinant of its dispersion matrix [11]. Minimizing the determinant of $D(\hat{\boldsymbol{\theta}})$, in turn maximizes the determinant of $M(\hat{\boldsymbol{\theta}})$, the information matrix. Both M and $D(\hat{\boldsymbol{\theta}})$ depend on the choice of the design ζ. The optimal design ζ^* for this problem would thus fulfill

$$\zeta^* = \arg\max_{\zeta} |M(\zeta, \hat{\boldsymbol{\theta}})| = \arg\min_{\zeta} |D(\zeta, \hat{\boldsymbol{\theta}})|. \tag{2.18}$$

This is called the D-Optimality, where the goal is to construct the design in a way that the determinant $|D(\hat{\boldsymbol{\theta}})|$ is minimized [11, 53]. For nonlinear parametrization, when equation 2.15 does not hold, calculating the information matrix is slightly more complicated. Assuming that $\eta(\boldsymbol{x}, \boldsymbol{\theta})$ is smooth around the true parameter values $\boldsymbol{\theta_0}$ and further that $\hat{\boldsymbol{\theta}}$ is located within this smooth region around $\boldsymbol{\theta_0}$, we can write the information matrix of the parameter estimates as follows [11]:

$$\begin{aligned} M(\hat{\boldsymbol{\theta}}) &= \sum_i^N w_i \boldsymbol{f}_\nabla(\boldsymbol{x}_i) \boldsymbol{f}_\nabla^T(\boldsymbol{x}_i) \\ with \quad \boldsymbol{f}_\nabla(\boldsymbol{x}_i) &= \nabla \eta(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}}) = \left[\frac{\partial \eta(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}})}{\partial \theta_0}, \frac{\partial \eta(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}})}{\partial \theta_1}, ..., \frac{\partial \eta(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}})}{\partial \theta_k} \right]^T \end{aligned} \tag{2.19}$$

Local Optimal Designs As discussed in the last section, constructing the information matrix for nonlinear parametrization is not entirely straight forward. In fact, the partial derivatives in equation 2.19 formally need to be evaluated at $\boldsymbol{\theta_0}$. This means that the optimal design for this problem depends on the true parameter values of the model, which, of course, are generally unknown. Thus, we can not construct a global D-optimal design for this case.

However, as was done in equation 2.19, we can construct approximations if we have some prior knowledge on the model parameters. For example, it might be possible to calculate an initial guess for $\hat{\boldsymbol{\theta}}_0$ based on some preliminary measurements. This kind of procedure is especially relevant in the context of sequential designs, as the parameter estimate is updated in every iteration. Designs for nonlinear parametrization, which are based on some approximation of $\hat{\boldsymbol{\theta}}_0$ are then called local optimal designs [11].

Sequential Design Fedorov proposed an algorithm for a sequential design for nonlinear parametrization[11]. As mentioned before, some initial experiments are necessary to obtain an estimate of the model parameters. The initial design of N points should be nondegenerate to allow for a single valued estimation of $\hat{\boldsymbol{\theta}}(N)$. The coordinates of the subsequent point $\boldsymbol{x}_{N+1}$ are then determined by solving the following optimization problem:

$$\max_{\boldsymbol{x}} \lambda(\boldsymbol{x}) d_N(\boldsymbol{x}) \quad with \quad d_N(\boldsymbol{x}) = \boldsymbol{f}_{\nabla,N}^T(\boldsymbol{x}) D(N) \boldsymbol{f}_{\nabla,N}(\boldsymbol{x}), \tag{2.20}$$

where we have replaced the weights w_i with the measurement efficiency function $\lambda(\boldsymbol{x})$. This is necessary, as we want to evaluate equation 2.20 at values $\boldsymbol{x}_{N+1}$ which are not contained in the initial design. Hence, also the variance b_{N+1}^2 is unknown. $\lambda(\boldsymbol{x})$ is an approximate function, which allows us to evaluate the efficiency of an measurement at any given $\boldsymbol{x}$. If no higher level information is available, a possible way of obtaining $\lambda(\boldsymbol{x})$ would be to fit an additional regression model to the observed variances of the initial measurements. The second term in the optimization, $d_N(\boldsymbol{x})$, corresponds to the approximate variance of $\eta(\boldsymbol{x}, \hat{\boldsymbol{\theta}})$ at the point $\boldsymbol{x}$. The dispersion matrix $D(N)$ is constructed from the information matrix of the N data points contained in the initial design according to

$$D^{-1}(N) = M(N) = \sum_{i}^{N} \lambda(\boldsymbol{x}_i) \boldsymbol{f}_{\nabla,N}(\boldsymbol{x}_i) \boldsymbol{f}_{\nabla,N}^T(\boldsymbol{x}_i). \tag{2.21}$$

The partial derivatives $\boldsymbol{f}_{\nabla,N}(\boldsymbol{x})$ are also evaluated at the current estimate of the model parameters $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(N)$. Iteratively adding new points to the design as dictated by equation 2.20, should yield a local optimal design in every step [11, 54].

2.4 Power Transformations

In statistical data analysis, power transformations are used to reduce the skew of distributions [55]. Especially in the context of linear regression modeling, such transformations can be of use. By introducing nonlinear transformations in a preprocessing step, often not only the quality of the fit will improve, but also the validity of the underlying assumptions [2]. For example, the functional relationship between two variables x_1 and x_2 may be highly nonlinear. Applying a power transformation to one (or both) quantities could unveil a simple linear relationship of the transformed variables. The transformation parameters can then also be chosen in way, such that the residual distribution of the model is close to a normal distribution. After all, normality of errors is one of the key assumptions in standard linear regression [33].

2.4.1 Tukey Ladder of Powers

Tukey and Mosteller [56] introduced the so called ladder of reexpressions in the context of exploratory data analysis [55]. They describe a very simple power transformation, which in the following will be referred to as the Tukey-transformation:

$$f_{Tukey}(x_i, \alpha_i) = \begin{cases} x_i^{\alpha_i} & if \quad \alpha_i \neq 0 \\ ln(x_i) & if \quad \alpha_i = 0 \end{cases} \tag{2.22}$$

The parameter α_i needs to be chosen based on the data at hand. What makes this kind of transformation very appealing is that it is able to represent many common functional relationships found in physics and chemistry. Some examples are given in here:

α	representation
-2	$1/x^2$
-1	$1/x$
-1/2	$1/\sqrt{x}$
-1/3	$1/\sqrt[3]{x}$
0	$\ln(x)$
1/3	$\sqrt[3]{x}$
1/2	$\sqrt{x}$
1	x
2	x^2

There is, however, one caveat to this transformation, which may become an issue if one would be to optimize α in some kind of algorithm. The ladder of reexpressions has a discontinuity for $\alpha = 0$. If the optimization of α can be restricted to values unequal 0 for some reason, the Tukey-transformation should still be applicable. These power transformations are usually applied to reduce skewness in the underlying data and bring it closer to being normally distributed. In the context of regression modeling, normally distributed residuals are desired as this is an indication of the adequacy of the model [2]. Based on the residual vector of a given model the transformation parameters α can be estimated.

2.4.2 Other power transformations: Box-Cox and Yeo-Johnon

Box and Cox introduced a slightly modified power transformation which solves the discontinuity problem for a transformation parameter of 0 by an additional shift and scaling factor[57]:

$$f_{Box-Cox}(x_i, \alpha_i) = \begin{cases} \frac{x_i^{\alpha_i}-1}{\alpha_i} & if \quad \alpha_i \neq 0 \\ ln(x_i) & if \quad \alpha_i = 0 \end{cases} \tag{2.23}$$

While this kind of transformation is now continuous at $\alpha_i = 0$ it suffers from the fact, that it is only applicable to positive values of x_i. Therefore, it has to be ensured that all input data is strictly positive before applying a Box-Cox transform. In practice, this is often not an issue, as a shift to positive numbers is a affine transformation, which will not change the shape of the distribution and hence have no effect on the optimal α. Nevertheless, Yeo and Johnson[58] came up with a generalized power transform, which does not have this requirement. The Yeo-Johnson transform can be applied to any real number and its transformation law reads as follows:

$$f_{Yeo-Johnson}(x_i, \alpha_i) = \begin{cases} \frac{(x_i+1)^{\alpha_i}-1}{\alpha_i} & if \quad x_i \geq 0 \quad and \quad \alpha_i \neq 0 \\ ln(x_i+1) & if \quad x_i \geq 0 \quad and \quad \alpha_i = 0 \\ \frac{-(-x_i+1)^{2-\alpha_i}-1}{2-\alpha_i} & if \quad x_i < 0 \quad and \quad \alpha_i \neq 2 \\ -ln(-x_i+1) & if \quad x_i < 0 \quad and \quad \alpha_i = 2 \end{cases} \tag{2.24}$$

2.5 The Probability Plot Correlation Coefficient

Filliben introduced the probability plot correlation coefficient (PPCC) normality test in 1975 [7]. A probability plot is a correlation plot between the ordered observations of a given data set and the corresponding theoretical quantiles drawn from a reference distribution. If one wants to check for the normality of the observations, the reference distribution is chosen to be a standard normal distribution. The PPCC is then defined as the pearson correlation

coefficient between these two vectors as given in equation 2.25.

$$PPCC = Corr(X, M) = \frac{\sum_{i=1}^{N} (X_i - \bar{X})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^{N} (X_i - \bar{X})^2 \sum_{i=1}^{N} (M_i - \bar{M})^2}}, \tag{2.25}$$

where X and M are the observations and theoretical quantiles with their respective mean values $\bar{X}$ and $\bar{M}$. The PPCC will result in values close to unity, if normality assumptions are valid for the underlying data, and smaller values for the case of different distributions. Figure 2.2 shows this behavior for samples drawn from a poisson, a uniform and a normal distribution.

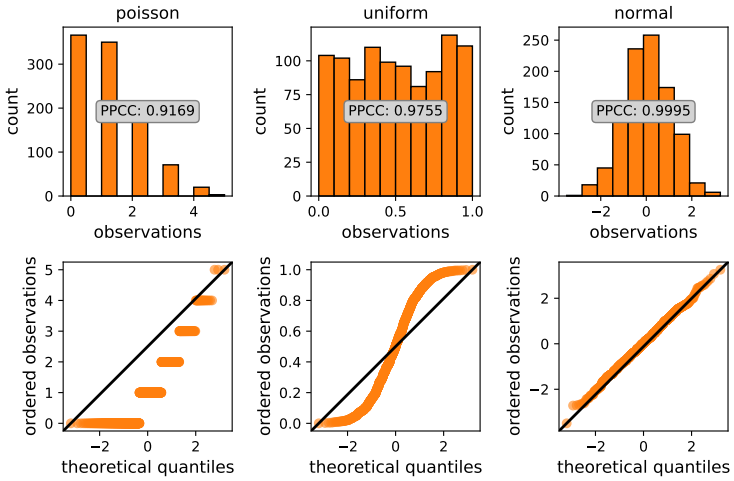


Figure 2.2: Probability plots for 1000 random samples drawn from different distributions. The corresponding PPCC values are given in the histograms.

In Figure 2.2 the sample size was chosen to be 1000 and indeed this sample size is limiting the inference from the PPCC. Figure 2.3 provides an estimate on the minimum sample size necessary to distinguish between different distributions. While the poisson data can clearly be identified as being non-normal, differentiating between uniform and normal may become an issue in the limit of small datasets. The PPCC values in Figure 2.3 are the average value of multiple runs per sample size. Therefore, in a single measurement the relative order of these distributions may actually be different. A soon as the sample sizes approaches roughly 100, however, the distributions should be accurately distinguishable.

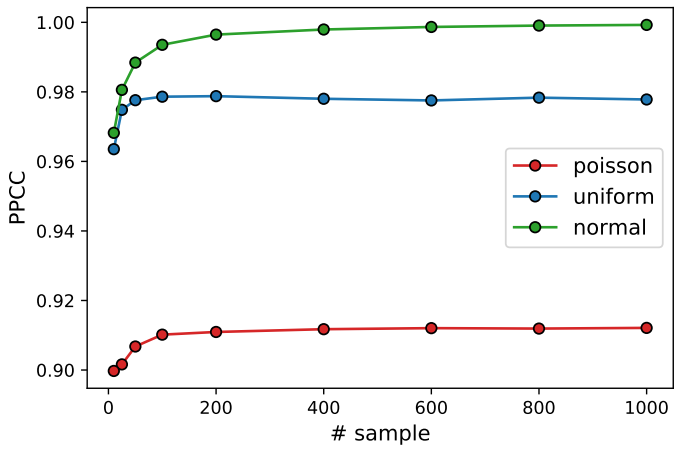


Figure 2.3: PPCC values for random samples of different size drawn from different distributions. The observed PPCC is itself averaged over 100 runs with fixed sample size.

2.6 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a numerical, gradient free, global optimization method [59]. It is considered a nature inspired algorithm and tries to exploit swarm intelligence in order to locate the minimum of a function. PSO enjoys great popularity in various fields, due to its conceptual simplicity and applicability to a wide range of optimization problems. Similar to genetic algorithms, PSO relies on a population (or swarm) of individuals to explore the search space at the same time. However, these individuals (called particles) locally interact with each other during the optimization, according to specified rules. In the simplest case, the PSO algorithm consists of the following steps[60]:

1. Initialize the positions ($x_i(t = 0)$) and displacements ($v_i(t = 0)$) of all N particles drawing random samples from a d-dimensional uniform distribution.

$$\begin{aligned} x_i(t=0) &= U(x_{min}, x_{max}, d) \\ v_i(t=0) &= U(v_{min}, v_{max}, d) \end{aligned} \tag{2.26}$$

2. Identify the current best position of the swarm (x^S_{min}) as well as the best position found for each particle up to this point ($x^{P_i}_{min}$)

3. Check if the convergence criterion is met. If yes, return x^S_{min}.

4. Update the positions and displacements according to:

$$\begin{aligned} v_i(t+1) &= \omega v_i(t) \\ &+ c_1 \cdot U(0,1,1) \cdot (x^{P_i}_{min}(t) - x_i(t)) \\ &+ c_2 \cdot U(0,1,1) \cdot (x^{S}_{min}(t) - x_i(t)) \end{aligned} \tag{2.27}$$

$$x_i(t+1) = x_i(t) + v_i(t+1)$$

   Go back to step 2.

The propagation rules in equation 2.27 describe the interaction of the swarm particles. There is a stochastic element contained in the particle displacement vectors, as two of the three contributions are scaled by random numbers between 0 and 1 drawn from a uniform distribution ($U(0,1,1)$). Further, some empirical parameters enter the equation. ω is related to the inertia of the particle trajectories as it dictates the contribution from the displacement of the previous step. The parameters c_1 and c_2 control the extent, to which the trajectory of the particle is biased towards its best know position or the best known position of the swarm, respectively. Thus, these parameters define the exploration-exploitation trade off in the optimization. In practice, the displacement vectors are usually bound to some finite range of values to avoid particles leaving the region of interest. Also, there are modifications to this algorithm, where the particles only communicate with a certain number of nearest neighbors[60].

2.7 Multi-Objective Optimization

The goal in an optimization problem is to minimize the value of some real valued function $f(\mathbf{x})$. This function is generally referred to as the objective function of the problem, while the $\mathbf{x}$ are called design variables. By varying these design variables we try to identify the optimal set of $\mathbf{x}$, such that the value of the objective function is optimized. There are various classes of optimization problems, with specialized algorithms for their respective solution. For an introduction to the topic the reader is referred to Polak[61]. In this work, we will deal with nonlinear optimization problems of continuous design variables with inequality constraints. Further, the number of objective variables is larger than one, meaning $f(\mathbf{x})$ itself is vector valued. In such cases, we speak of multi-objective optimization, as we want to find the set of $\mathbf{x}$ which optimizes all objectives in the best possible way[62].

2.7.1 The Pareto Front

While in scalar, single objective problems there is usually one global optimum, this is not necessarily the case in a multi-objective optimization. If two or more objectives are interfering with each over, at some point it might not be possible to decrease the value of objective f_1 while not increasing the value of another objective f_2. This leads to a multitude of possible tradeoffs between these two objectives. To reduce the number of these tradeoff points to the necessary minimum, we distinguish between dominated and non-dominated solutions. A solution x_1 is dominating another solution x_2 if all objective values for x_1 are at least as good as for x_2 and in at least one objective x_1 is better than x_2. This relation is schematically shown in Figure 2.4. If x_1 dominates x_2, x_2 must be located in the red area and have worse objective values in both f_1 and f_2. On the other hand if x_2 is located in the green area, we can say, that it dominates x_1. For the other two quadrants, x_2 would be better in one of the objectives and worse in the other. Therefore, we can not differentiate, whether x_1 or x_2 is more optimal.

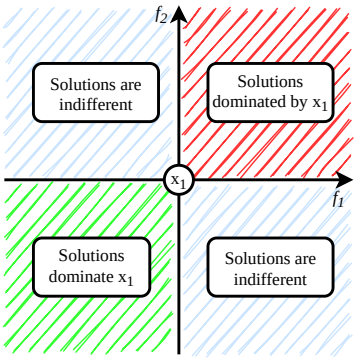


Figure 2.4: Illustration of the 2D objective space spanned by f_1 and f_2. The different areas indicate the dominating relations between an arbitrary solution x_1 and other solutions in the four quadrants. The illustration is based on Collete[62].

The non-dominated solutions are also referred to as optimal in the Pareto sense (or Pareto optimal). If we reduce the number of tradeoff solutions in a multi-objective problem to only those, which are non-dominated, we end up with the so called Pareto Front (PF). This PF is a subspace of the objective space, whose dimensionality is one lower compared to the objective space. This means that for a problem with two objectives, the PF would be a 1D curve, for three objectives the PF is a 2D surface and so on. As all solutions located on the PF are Pareto optimal, they are all indifferent with respect to each other. Hence, we can not decide which solution to pick without additional information. A schematic of the PF in a 2D objective space is shown in Figure 2.5.

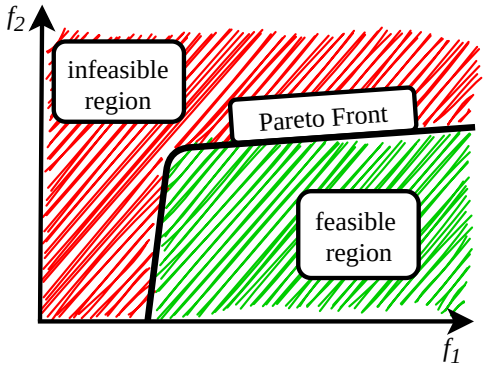


Figure 2.5: Illustration of the Pareto Front (PF) in a 2D objective space spanned by f_1 and f_2. The PF separates the feasible from the infeasible set of solutions. In this case the PF is a 1D curve, as the overall objective space is two dimensional. The shape of this PF is typical for a problem, in which f_1 needs to be minimized, while f_2 is supposed to be maximal.

The question is now, how to determine this PF in form of an optimization problem.

2.7.2 Evolutionary Multi-Objective Optimization (EMO)

Evolutionary optimization techniques are iterative algorithms designed to mimic the evolutionary processes encountered in nature. They operate on a population of candidate solutions and use operations like selection, crossover and mutation to update this population towards an optimal solution. Such an optimization approach is very well suited for determining the PF in a multi-objective problem, as we are already working on a population of multiple solutions. Further, incorporating stochastic events like mutations in the optimization process, allows the algorithm to handle multiple optima more easily. Also, evolutionary algorithms do not rely on gradients in order to update the solutions making them suited for complex nonlinear problems. In general, evolutionary algorithms consist of the following operators [63]:

Initialization In a first step, an initial population is created by sampling the available design space of the problem. Usually, upper and lower bounds are defined by the optimization task. In principle, any kind of sampling algorithm may be applied here. Lacking additional information, however, the initial population is often drawn randomly.

Evaluation In the second step, for every individuum in the population the fitness needs to be determined. This means, that the cost function of the optimization problem is evaluated for every candidate solution. Both the violation of constraints and the objective values determine the fitness ranking within the population.

Selection Here, individuals are selected from the population to form the mating pool for the next generation. A common variant is the tournament selection, which picks two individuals at random and select only the "fitter" of the two for the mating pool.

Variation In this step, parent individuals are selected from the mating pool and used to create new individuals by exchanging information. The variation of the population can also incorporate multiple processes like mutation and crossover. Such additional stochastic steps can help to escape local optima and diversify the population.

Elitism During the elitism operation, both, the population of the last generation and of the current generation are combined. From these joined populations the individuals for the next generation are chosen. This enables the algorithm to keep good solutions from previous generations and it could be shown, that such an elitism is required to ensure a non-degrading performance of the algorithm [8, 64].

Termination A common termination criterion is simply a maximum number of either iterations (generations) or objective function evaluations. In EMO alternatively also measures corresponding to the convergence of the PF can be applied. E.g. the change in hypervolume occupied by the PF in objective space can be a reasonable termination criterion [65, 66].

The goal of an EMO is to find a set of solutions, which resemble the PF of the multi-objective problem as closely as possible. Ultimately, one Pareto optimal solution needs to be selected from the PF. Therefore, we want the approximated PF from the EMO to not only contain non-dominated solutions, but also as diverse trade offs as possible. A well established EMO algorithm is Non-dominated Sorting Genetic Algorithm (NSGA-II) [8]. It uses elitist principles as well as mechanisms to emphasize both non-dominated and maximally diverse solutions. In every iteration, the current parent and offspring populations are combined. Then, this larger set of candidate solutions is sorted into non-domination fronts and only the top half of the individuals is kept to restore the original population count. If only some individuals from a non-domination front can be selected, the crowding distance sorting is applied. The crowding distance corresponds to the volume an individual occupies in objective space in which no other solution is located. By selecting those individuals with the maximum crowding distance, the population of the next generation will be as diverse as possible. This procedure is schematically illustrated in Figure 2.6.

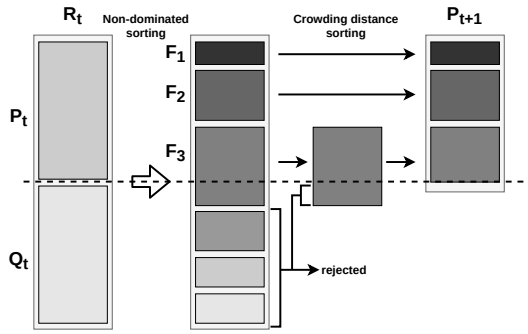


Figure 2.6: Schematic of the non-dominated and crowding distance sorting within NSGA-II. P_t and Q_t are the current parent and offspring populations, respectively. Those are combined and the resulting population is sorted into non-domination fronts (F_1, F_2, F_3, ...) of different classes. According to this sorting, the top half of the population is kept to form the new population P_{t+1}. In the highly probable case that not all individuals of a non-domination front can be kept, the best individuals according to the crowding distance sorting are selected. The illustration is based on Branke[63].

2.8 Quasi Monte Carlo Sampling

Quasi Monte Carlo (QMC) sampling is an approach to efficiently cover a high dimensional space for multivariate integration. To this end, QMC samples rely on so called low-discrepancy sequences. With every given number of points, a QMC sample tries to represent the n-dimensional uniform distribution as closely as possible. In contrast to random sampling, there is no requirement for the points to be independent. Hence, QMC samples can also be constructed using deterministic sequences. A major construction principle of such sequences is the so called full projection regularity. For a sequence P_n with m points in n dimensions to be fully projection regular, its projections on to point sets $P_{k<n}$ need also to contain m distinct points [10]. The design of QMC samples aims at a faster convergence compared to standard Monte-Carlo methods, given that the function to integrate is sufficiently smooth within the region of interest. There are two different types of QMC methods: the open and the closed type. An open type QMC sample is constructed from an infinite sequence of points, which are also referred to as digital sequences. This comes with the advantage that, to increase the number of points in the sample, simply the subsequent points in the sequence need to be measured. Closed type QMC samples (or digital nets) on the other hand, usually result in a completely new point set, if the number of measurements is changed [10, 67]. Deterministic QMC sequences are a convenient way of sampling a multidimensional space, if no

prior knowledge is available. Given a number of function evaluations and dimensions of the sample space, a QMC set can be generated and simply scaled to the corresponding ranges. Two well known QMC sequences are the Halton [68] sequence and the Sobol [69] sequence, which both count among the open type digital sequences. The first 64 points of these two sequences for the two dimensional case are illustrated in Fig. 2.7.

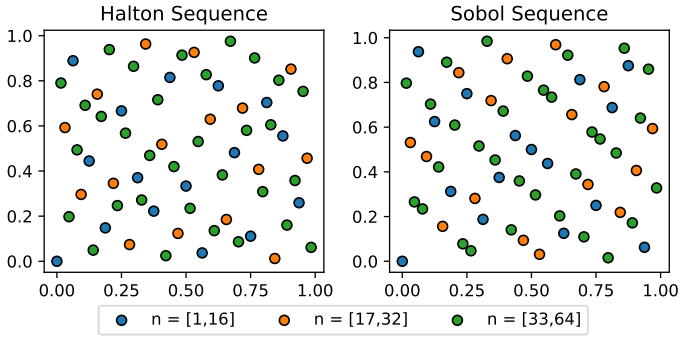


Figure 2.7: Examples for QMC point sets chosen as a two dimensional Halton or Sobol sequence, respectively. The point sets each consist of n=64 samples. The colors indicate different portions of the sequences.

2.9 Clustering Analysis

2.9.1 Proximity Based Clustering: K-Means

When evaluating large data sets, cluster analysis is a convenient way of reducing the amount of data points while containing as much information as possible [70]. The idea is to split the data into subsets properly represented by one cluster center, the so called centroid. The datapoints are sorted into clusters according to some similarity criterion. There are many different algorithms available for performing such a task, one of them being *Kmeans++*.

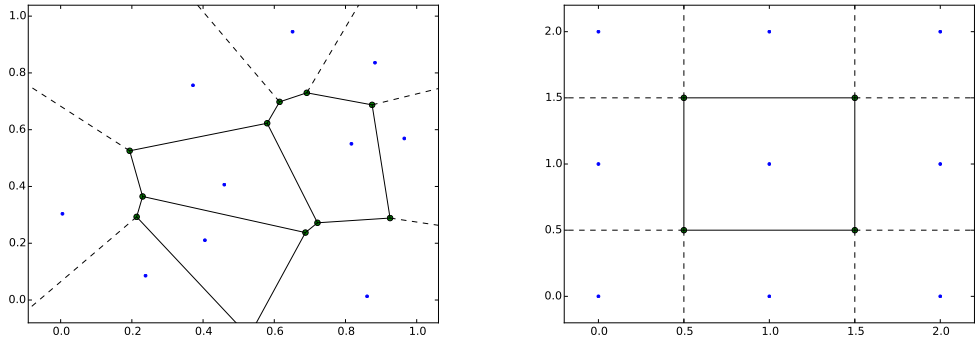
Lloyd Algorithm The *Lloyd* algorithm is an algorithm used to create centroidal *Voronoi* tesselations (CVT) of data sets. A Voronoi tesselation is a splitting of a given set of points into subsets referred to as *Voronoi* regions. Each *Voronoi* region is defined by a generating point according to equation 2.28.

$$\hat{V}_i = \{\mathbf{x} \in \Omega \quad | \quad | \mathbf{x} - \mathbf{z}_i | < | \mathbf{x} - \mathbf{z}_j | \quad \text{for } j = 1, ..., k, j \neq i\}, \tag{2.28}$$

where $\mathbf{z}_i$ is the generator, Ω is a general open set $\Omega \subseteq \mathbb{R}^N$ and $\mathbf{x}$ are the points contained in this region [71]. These are all points of the original set closer to its generator than to the generator of every other region. The tesselation is referred to as centroidal if the generating point of a region coincides with its respective mass center $\mathbf{z}_i^*$. Given a density function ρ defined on Ω, the mass center can be defined according to equation 2.29 [72].

$$\mathbf{z}_i^* = \frac{\int_{V_i} \mathbf{y}\rho(\mathbf{y})d\mathbf{y}}{\int_{V_i} \rho(\mathbf{y})d\mathbf{y}} \tag{2.29}$$

From figure 2.8(a), showing the *Voronoi* diagram for a set of randomly chosen points, it is obvious that the generator of a region does not have to be its mass center. In the case of a rectangular lattice however, shown in figure 2.8(b) the tesselation with respect to these points directly yields a centroidal one. Thus in order to create a CVT for a generic data set an iterative approach is necessary. The *Lloyd* algorithm is such a method. Given the previously defined open set Ω as well as an initial set of generators $\{z_i\}_{i=1}^k$, in the first step this algorithm creates a *Voronoi* tesselation. Next, the center of mass $\mathbf{z}_i^*$ is calculated for every region V_i. These centers are then used as the generators for the *Voronoi* tesselation of the following iteration. This procedure is repeated until a convergence criterion is met, for



(a) (b)

Figure 2.8: 2D *Voronoi* diagrams for (a) a set of random points and (b) points on a rectangular lattice; the tesselation was performed according to the *Euclidian* distance of the data points.

example the sum of distances for all points with respect to their centroid reaches a certain value [72]. As with most iterative algorithms, choosing the optimal initial generators has a major impact on the efficiency of the tesselation process.

Kmeans++ Algorithm Creating a CVT of a data set basically is a cluster analysis. Given a number of cluster centers, the goal is to minimize the total sum of distances between the points and their respective centers. This is analog to the *Lloyd* algorithm with the according convergence criterion. The difficult task is to choose the initial cluster centers, such that the algorithm converges quickly and yields a reasonable segmentation. Within the standard *Kmeans* algorithm initial centers are chosen arbitrarily. By introducing a more sophisticated seeding procedure the *Kmeans++* algorithm tries to improve upon this completely random process: At first, one (of k) center c_1 is chosen uniformly random from the data set Ω. The next center is obtained by choosing a $\boldsymbol{x} \in \Omega$ according to the probability distribution given in 2.30.

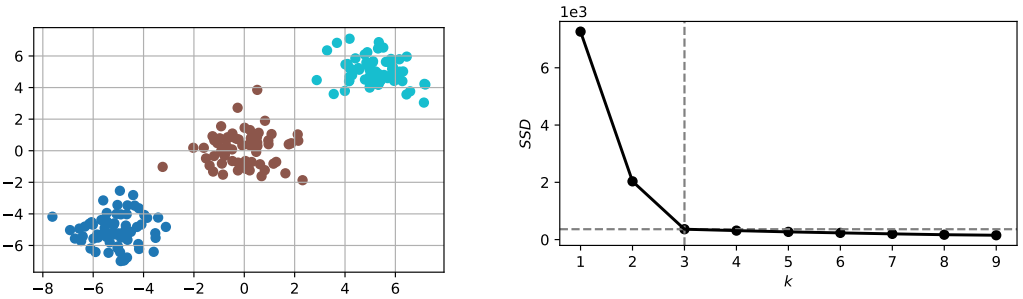
$$P(\boldsymbol{x}) = \frac{D(\boldsymbol{x})^2}{\sum_{\boldsymbol{x}\in\Omega} D(\boldsymbol{x})^2} \quad , \tag{2.30}$$

where $D(\boldsymbol{x})$ refers to the shortest distance between any data point to a center already chosen. This procedure is repeated until k centers are obtained. These form the initial guess for the standard *Kmeans* algorithm.

Scree-Plot One drawback of the *Kmeans* algorithm is the requirement for manually defining k, the number of clusters. In general, k is unknown and it may not always be possible to estimate it based on higher level information on the problem at hand. Several different methods can be applied to determine k in such cases. One conceptually simple option is the so called Ellbow method[73]. In order to determine the optimal k_{opt}, we perform a clustering of the data set for different values of k and plot the resulting clustering errors. In the context of *Kmeans*, the clustering error (sometimes also referred to as inertia) is usually the sum of squared distances (SSD) for all points w.r.t. to the corresponding centroid as given in equation 2.31.

$$SSD = \sum_{c=1}^{k} \sum_{m\in I_c}^{N_c} |\boldsymbol{x}_m - \boldsymbol{z}_c|^2 \quad , \tag{2.31}$$

where c runs over all clusters with their respective centroids $\boldsymbol{z_c}$ and m runs over the point set I_c assigned to each cluster. The resulting plot is called the scree plot and shows a continuously decreasing SSD with increasing k. While initially, the increments in SSD are large, the SSD values usually quickly converge towards 0, when k approaches the number of data points. In case a meaningful splitting of the data is possible, the scree plot should show an ellbow (or sometimes knee) point like indicated in figure 2.9(b). At this point, the scree curve



(a) (b)

Figure 2.9: (a) A hypothetical 2D data set composed of three distinct clusters and (b) the corresponding scree plot of the clustering error (SSD) versus the number of clusters k. The dashed lines indicate k_{opt} coinciding with the knee point of the scree curve.

shows the maximum curvature. This point corresponds to k_{opt}. The simplest way of identifying this point is by visual inspection, however, automatizing this step is possible, e.g. by numerically approximating the curvature [74].

2.10 Support Vector Classification

2.10.1 Linear Support Vector Machines

The support vector machine (SVM) is a statistical learning model applicable to the classification and regression of data [9]. It is based on the principle of separating hyperplanes, splitting the data set into multiple classes. Dealing with a space of dimension N, a hyperplane is a $N-1$ dimensional subspace, which can be used to separate the original space into multiple regions. Hence, in 3D space, the term hyperplane would refer to a 2D plane, in 4D space to a 3D subspace and so on. Given a set of points consisting of two classes, the optimal separating hyperplane not only separates the two subsets but also maximizes the distance from its surface to the points of either class, the so called margin. Identifying such a hyperplane can be formulated in terms of an optimization problem. However, for this optimization to find a feasible solution, the data sets needs to be linearly separable. To overcome this problem, we can use a SVM. We can set up the SVM as a hyperplane based classification model, even if the requirement of linear separability in the training data is not given. In the case of overlapping sample distributions, the SVM allows for a certain degree of misclassification by introducing a penalty term. The optimization problem for the linear support vector classifier (SVC) is the following [70]:

$$\min_{\beta_0,\beta} \frac{1}{2}||\boldsymbol{\beta}||^2 + C\sum_{i=1}^{N_p}\xi_i \quad subject\ to \quad \xi_i \geq 0, \quad y_i(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{\beta_0}) \geq 1-\xi_i \quad \forall i, \tag{2.32}$$

where β and β_0 are the coefficients of the hyperplane defined by the equation $(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{\beta_0}) = 0$, C is the tuning parameter of the penalty and ξ_i is a slack variable corresponding to the distance to the correct margin boundary for every point located within the margin. As the margin width is defined as $M = \frac{1}{||\boldsymbol{\beta}||}$ the minimization in equation 2.32 leads to a maximized margin. The magnitude of the penalty coefficient C controls the contribution due to misclassification. In the limit of $C = \infty$, equation 2.32 would correspond to the optimal separating hyperplane problem. A visualization of a two class SVM is shown in figure 2.10.

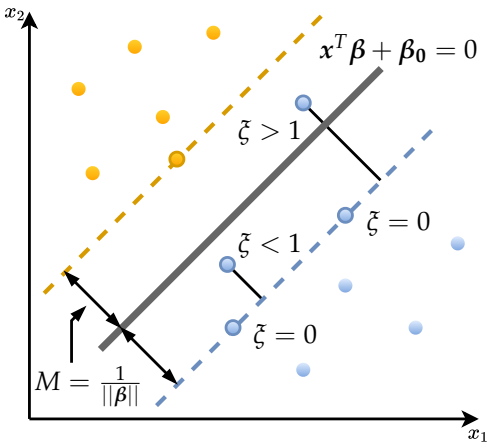


Figure 2.10: 2D example of a two class SVM. The solid gray line is the separating plane between the two classes (orange and blue data points). Data points surrounded by a circle are support vectors. The dashed lines indicate the margin boundaries. The illustration is based on Bishop[75].

Having determined the hyperplane parameters, we can define the decision function $G(x)$ to classify new observations, not contained in the original training data:

$$G(\boldsymbol{x}) = sgn(\boldsymbol{x}^T\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}_{\mathbf{0}}), \tag{2.33}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$ are the solutions of the optimization problem in equation 2.32.

2.10.2 Estimating Membership Probabilities: Platt Scaling

Using the decision function of a SVC, we can predict the class of an unknown data point $\boldsymbol{x}_i$. Additionally, we can estimate its score f_i, which is proportional to the signed distance of $\boldsymbol{x}_i$ to the decision boundary (the separating hyperplane). However, in certain cases it might be useful to have a calibrated probability of the point $\boldsymbol{x}_i$ being assigned to the class y, a posterior probability. For the two class problem, Platt introduced a method to estimate this posterior probability in 1999 [76]. Therein, a parametric sigmoid function is assumed as a model for the probabilities:

$$P(y = 1|f) = \frac{1}{1 + exp(Af + B)}, \tag{2.34}$$

where $y = 1$ indicates the membership in class 1 and A and B are scalar parameters. These parameters can be estimated from maximum likelihood and the resulting optimization problem can be written as follows:

$$\begin{aligned} \min_{A,B} \left[-\sum_{i=1}^{N} t_i \log(p_i) + (1 - t_i)\log(1 - p_i) \right] \quad , \\ t_i = \frac{y_i + 1}{2} \quad , \\ p_i = \frac{1}{1 + exp(Af_i + B)} \quad . \end{aligned} \tag{2.35}$$

To avoid a bias of the probability calibration, this fit is not performed on the same training data as the fit of the original SVC. Instead, a n-fold cross validation is performed and the SVC scores f_i are determined on the test set of each fold. For multi class problems, Wu et al. proposed an extension of this formalism by combining the comparisons of all two class sub problems [77]. Assuming that all pairwise probability estimates r_{ij} are available, the multi

class problem for k different classes is the following:

$$\min_{p} \sum_{i=1}^{k} \sum_{j \neq i}^{k-1} (r_{ji} p_i - r_{ij} p_j)^2 \quad subject\ to \quad \sum_{i=1}^{k} p_i = 1, \quad p_i \geq 0, \forall i. \tag{2.36}$$

2.11 Chemical Reaction Systems

Four chemical processes are exemplarily studied in this work to illustrate the suggested algorithm. The first two reaction networks, the classical gas phase HBr reaction and a generic catalytic cycle, can be attributed to homogeneous reactions. The latter two examples represent heterogeneous processes. Microkinetic models for the catalytic oxidation of carbon monoxide both, over a Pt catalyst and over a RuO_2 surface are investigated.

2.11.1 Hydrogen Bromine Reaction

Hydrogen (H_2) and bromine (Br_2) in gas phase can undergo a reaction to form hydrogenbromide (HBr):

$$H_2 + Br_2 \rightarrow 2HBr \quad . \tag{2.37}$$

This reaction is a classic example, where the apparent initial reaction orders are not obvious from the stoichiometric coefficients in the formal reaction equation. This indicates that a more complex microkinetic mechanism takes place. Bodenstein and Lind [78] described an effective rate law for the formation of HBr as

$$\frac{d[HBr]}{dt} \propto [H_2] \cdot [Br_2]^{1/2} \quad . \tag{2.38}$$

This law was found to be valid in the early stages of the reaction in a regime, where $[HBr] << [Br_2]$. It corresponds to an effective reaction order of 1.5 which could later be explained by the, now well established, radical reaction mechanism [79]. The reaction mechanism and rate constants for the microkinetic model have been taken from experimental kinetic studies by Levy[80] and Cooley and Anderson[81]. The detailed mechanism is given in Table 2.1.

Dissociation	$Br_2 \xrightarrow{k_1} 2Br\cdot$
and	$2Br\cdot \xrightarrow{k_2} Br_2$
Recombination	$H_2 \xrightarrow{k_3} 2H\cdot$
	$2H\cdot \xrightarrow{k_4} H_2$
	$H\cdot + Br\cdot \xrightarrow{k_5} HBr$
	$HBr \xrightarrow{k_6} H\cdot + Br\cdot$
Chain	$Br\cdot + H_2 \xrightarrow{k_7} HBr + H\cdot$
Propagation	$Br\cdot + HBr \xrightarrow{k_8} Br_2 + H\cdot$
	$H\cdot + Br_2 \xrightarrow{k_9} HBr + Br\cdot$
	$H\cdot + HBr \xrightarrow{k_{10}} H_2 + Br\cdot$

Table 2.1: Elementary steps of the mechanism in the HBr model.

Table 2.2: Kinetic parameters for the HBr model.

reaction	prefactor [a.u.]	barrier [kcal/mol]	rate constants @ 850K [a.u.]
k_1	$7.18{\cdot}10^{12}$ * T	45.23	$1.43{\cdot}10^{4}$
k_2	$5.70{\cdot}10^{15}$	0.00	$5.70{\cdot}10^{15}$
k_3	$2.00{\cdot}10^{13}$ * T	102.48	$7.61{\cdot}10^{-11}$
k_4	$1.10{\cdot}10^{16}$	0.00	$1.10{\cdot}10^{16}$
k_5	$9.00{\cdot}10^{15}$	0.01	$8.95{\cdot}10^{15}$
k_6	$5.95{\cdot}10^{12}$ * T	85.85	$4.27{\cdot}10^{-7}$
k_7	$8.05{\cdot}10^{10}$ * T	17.70	$1.93{\cdot}10^{9}$
k_8	$9.31{\cdot}10^{10}$ * T	41.70	$1.50{\cdot}10^{3}$
k_9	$2.59{\cdot}10^{11}$ * T	1.10	$1.15{\cdot}10^{14}$
k_{10}	$3.08{\cdot}10^{10}$ * T	1.10	$1.37{\cdot}10^{13}$

All barriers and prefactors for the elementary processes are given in Table 2.2.

2.11.2 Catalytic Cycle

The second reaction process consists of a simple model catalytic cycle. Two reactants A and B form a complex ($*$AB) with the catalyst ($*$) in a pre-equilibration step. A subsequent reaction forms the reaction product P and restores the catalyst's initial state. Reaction mechanisms of this type are frequently encountered in homogeneous catalysis and enzymatic reactions. The reaction equations are given in Fig. 2.11. This specific reaction model is based on work by Pollice [82].

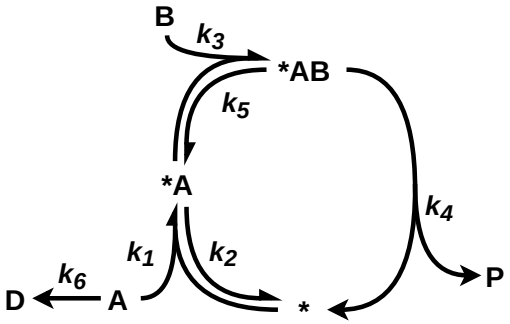


Figure 2.11: Reaction network of the catalytic cycle model. Reactants A and B form an activated complex with a catalyst (*), which reacts to form the reaction product P in a subsequent step. The k_i are rate constants for the respective elementary steps.

We differentiate between two cases: the undisturbed catalytic reaction and the cycle with a side reaction (k_6) leading to a removal of reactant A, e.g. by decomposition. Tables 2.3 and 2.4 show the initial homogeneous feed composition and the rate constants of the catalytic cycle model, respectively.

Table 2.3: Feed concentrations for the catalytic cycle model.

species	concentration [mol/l]
$[A]_0$	1.0
$[B]_0$	1.0
$[*]_0$	0.01
$[P]_0$	0.0
$[D]_0$	0.0
$[*A]_0$	0.0
$[*AB]_0$	0.0

Table 2.4: Rate constants for the catalytic cycle model.

reaction	prefactor [a.u.]	barrier [kcal/mol]	rate constants @ 700K [a.u.]
k_1	1.0	2.0	$2.37{\cdot}10^{-1}$
k_2	10.0	1.0	4.87
k_3	1.0	2.0	$2.37{\cdot}10^{-1}$
k_4	10.0	6.0	$1.34{\cdot}10^{-1}$
k_5	1.0	1.50	$3.40{\cdot}10^{-1}$
k_6	$1.0{\cdot}10^3$	5.00	$2.75{\cdot}10^{-5}$

2.11.3 CO oxidation on Pt

The catalytic oxidation of carbon monoxide to carbon dioxide is a classical example for a reaction following a Langmuir-Hinshelwood mechanism [83]:

$$\begin{aligned} \mathrm{CO} + * &\rightleftharpoons \mathrm{*CO} \\ \mathrm{O_2} + 2* &\rightarrow 2\,\mathrm{*O} \\ \mathrm{*O} + \mathrm{*CO} &\rightarrow \mathrm{CO_2} + 2* \end{aligned} \tag{2.39}$$

The oxidation of CO exhibits several interesting kinetic phenomena. For example, CO oxidation on a Pt catalyst is known for its oscillatory kinetics. The simple irreversible kinetic model used in this work is based on kinetic equations by Ertl and coworkers [84]. The detailed mechanism contains the following elementary steps:

$$\begin{aligned} \mathrm{O_2} + 2* &\xrightarrow{k_{ad,\mathrm{O_2}}} 2\mathrm{O}_{ad} \\ \mathrm{CO} + * &\xrightarrow{k_{ad,\mathrm{CO}}} \mathrm{CO}_{ad} \\ \mathrm{CO}_{ad} &\xrightarrow{k_{des,\mathrm{CO}}} \mathrm{CO} + * \\ \mathrm{CO}_{ad} + \mathrm{O}_{ad} &\xrightarrow{k_{reac}} \mathrm{CO_2} + 2* \\ \mathrm{O}_{ad} + \mathrm{Pt} &\xrightarrow{k_{oxid}} \mathrm{PtO}_{ad} \\ \mathrm{CO}_{ad} + \mathrm{PtO}_{ad} &\xrightarrow{k_{red}} \mathrm{CO_2} + \mathrm{Pt} + 2* \end{aligned} \tag{2.40}$$

Kinetic parameters and rate constants are given in Table 2.5.

Table 2.5: Rate constants for the CO oxidation model.

reaction	prefactor [a.u.]	barrier [kcal/mol]	rate constants [a.u.] @ 535 K
k_{ad,O_2}	$1.25 \cdot 10^4$	1.0	$4.88 \cdot 10^3$
$k_{ad,CO}$	$1.0 \cdot 10^3$	0.0	$1.0 \cdot 10^3$
$k_{des,CO}$	$2.70 \cdot 10^{10}$	20.0	$1.83 \cdot 10^2$
k_{reac}	$9.40 \cdot 10^8$	10.0	$7.73 \cdot 10^4$
k_{oxid}	$2.80 \cdot 10^{-1}$	1.0	$1.09 \cdot 10^{-1}$
k_{red}	$6.10 \cdot 10^1$	10.0	$5.0 \cdot 10^{-3}$

The analysis in this work always deals with the initial reaction regime, meaning very short residence times. For this reason, the surface oxidation does not play a significant role in the reaction kinetics.

2.11.4 CO Oxidation on RuO_2

The CO oxidation on RuO_2 is an extensively studied catalytic reaction system [12, 85–89]. In this work, we use a mean field MKM implemented as described by Temel et al. [12]. The model is propagated into a quasi stationary state by numerically integrating the corresponding system of differential equations. The space of physical parameters for this model is spanned by the partial pressures of O_2 and CO as well as temperature. The original authors[12] show, that in contrast to a kinetic Monte Carlo (kMC) model, the mean field model does not accurately describe the kinetics of this reaction system. However, it is also stated that the surface phase diagram as well as the sharp peak in turn over frequency (TOF) can be reproduced using the mean field model. The elementary reaction steps in the mean field MKM model involve two different adsorption sites on the RuO_2 (110) facet, the twofold coordinated bridge (br) and the fourfold coordinated (cus) site. An illustration of the position of these sites is given in Figure 2.12.

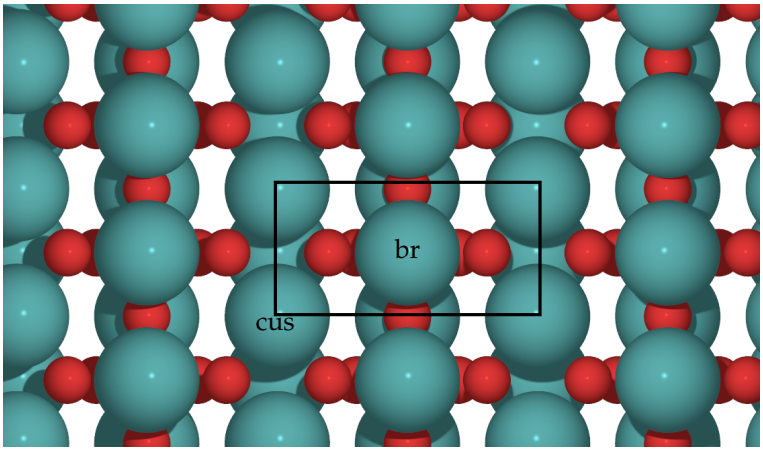


Figure 2.12: Top view of the RuO_2 (110) surface structure. The box indicates the surface unit cell and the two adsorption sites, the bridge site and the fourfold coordinated cus site are highlighted [12].

The following reaction steps are included:

$$
\begin{aligned}
\mathrm{CO_{br}} &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} CO_{gas} + *_{\mathrm{br}} \\
\mathrm{CO_{cus}} &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} CO_{gas} + *_{\mathrm{cus}} \\
2\,\mathrm{O_{br}} &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} O_{2,gas} + 2*_{\mathrm{br}} \\
2\,\mathrm{O_{cus}} &\underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} O_{2,gas} + 2*_{\mathrm{cus}} \\
\mathrm{O_{br}} + \mathrm{O_{cus}} &\underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} O_{2,gas} + *_{\mathrm{br}} + *_{\mathrm{cus}} \\
\mathrm{CO_{br}} + \mathrm{O_{cus}} &\xrightarrow{k_6} CO_{2,gas} + *_{\mathrm{br}} + *_{\mathrm{cus}} \\
\mathrm{CO_{br}} + \mathrm{O_{br}} &\xrightarrow{k_7} CO_{2,gas} + 2*_{\mathrm{br}} \\
\mathrm{CO_{cus}} + \mathrm{O_{cus}} &\xrightarrow{k_8} CO_{2,gas} + 2*_{\mathrm{cus}} \\
\mathrm{CO_{cus}} + \mathrm{O_{br}} &\xrightarrow{k_9} CO_{2,gas} + *_{\mathrm{br}} + *_{\mathrm{cus}}
\end{aligned}
\tag{2.41}
$$

The rate constants for all listed elementary steps are given in Table 2.6.

Table 2.6: Rate constants for the CO oxidation model on RuO_2 for two different conditions: T = 600 K, p_{CO} = 7.0 atm, p_{O_2} = 1.0 atm and T = 350 K, p_{CO} = $4.0{\cdot}10^{-10}$ atm, p_{O_2} = $1.0{\cdot}10^{-10}$ atm [12].

reaction	rate constants @ 600 K [a.u.]	rate constants @ 300 K [a.u.]
k_1	$2.80{\cdot}10^{4}$	$1.40{\cdot}10^{-6}$
k_{-1}	$7.20{\cdot}10^{8}$	$5.00{\cdot}10^{-2}$
k_2	$9.20{\cdot}10^{6}$	$2.90{\cdot}10^{-2}$
k_{-2}	$7.20{\cdot}10^{8}$	$5.00{\cdot}10^{-2}$
k_3	$4.10{\cdot}10^{-21}$	$2.00{\cdot}10^{-49}$
k_{-3}	$9.70{\cdot}10^{7}$	$1.00{\cdot}10^{-2}$
k_4	$2.80{\cdot}10^{1}$	$5.50{\cdot}10^{-12}$
k_{-4}	$9.70{\cdot}10^{7}$	$1.00{\cdot}10^{-2}$
k_5	$3.40{\cdot}10^{-10}$	$1.10{\cdot}10^{-30}$
k_{-5}	$9.70{\cdot}10^{7}$	$1.00{\cdot}10^{-2}$
k_6	$1.20{\cdot}10^{6}$	$1.10{\cdot}10^{1}$
k_7	1.60	$9.20{\cdot}10^{-10}$
k_8	$1.70{\cdot}10^{5}$	0.40
k_9	$5.20{\cdot}10^{2}$	$1.90{\cdot}10^{-5}$

Chapter 3

Results

"Komm hol das LASSO raus..." - Julius H.

3.1 Robust Sparse Polynomial Regression in Transformed Variables

The focus of this chapter is on the description of a novel kinetic modeling approach aimed at identifying effective rate laws. As established in section 2.1, reaction networks can be described by systems of rate equations (cf. equation 2.2). Looking at the functional form of such rate equations, we can see that kinetic rate laws generally resemble sums over products of the terms $[X_m]^{s_{jm}}$. Such products can also be thought of as interactions terms in a polynomial function. Restricting ourselves to effective local models, we can interpret kinetic rate laws as sparse polynomial functions in terms of the reactant concentration taken to their respective partial reaction order. This mathematical structure of the kinetic rate laws heavily influenced the selection of methods, incorporated in the proposed modeling algorithm, which tries to identify a robust and data efficient mean field representation of the underlying kinetics from empirical data. The first section of this chapter will describe the general form of the modeling approach, followed by several examples from classical kinetics as a proof of concept. Ultimately, we apply the method to a catalytic reaction to highlight its capability to differentiate between regimes of distinct effective kinetics. The work presented in this chapter has been published in the Chemical Engineering Journal under the CC-BY license [1].

3.1.1 Introduction

Detailed knowledge about reaction kinetics for large and industrially relevant reaction networks is the basis for a rational design and efficient operation of chemical reactors. However, when studying a new catalytic system the underlying mechanism is in general unknown. Concentration profiles obtained at different reaction conditions contain valuable information about the (inter-) dependence of reactants and products. Extracting these dependencies to gain a deeper understanding of the catalytic mechanism is of general importance in heterogeneous as well as in homogeneous thermal catalysis, on which we will focus here without loss of generality. As illustrated in Figure 3.1 the central observable in form of a concentration profile is independent of the catalytic process at hand. In homogeneous catalysis, reactions are usually performed in a batch reactor. Probing the reaction mixture at different points in time and quenching the respective samples directly yields a concentration profile. When studying heterogeneous catalysts in a gas flow reactor the same information is contained in profiles of concentrations versus residence time in the reactor bed. The residence time can be adjusted by either modulating the gas flow through the reactor or by taking samples at different points along the reactor tube.

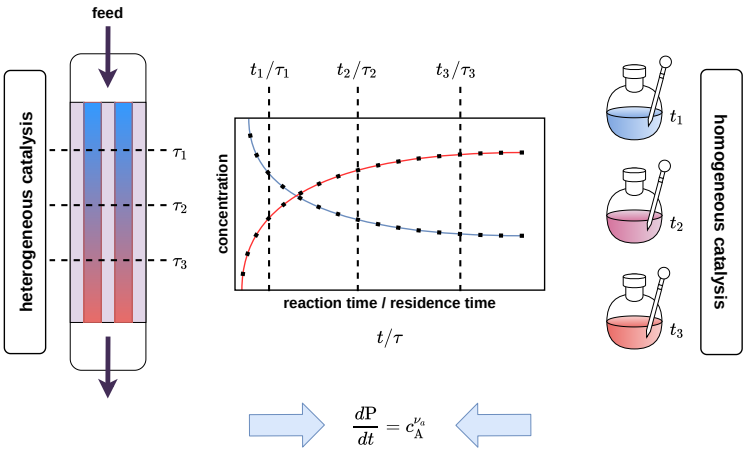


Figure 3.1: Concentration-time profiles (e.g. from flow (left) or batch (right) measurements) as central quantity in the kinetic analysis of catalytic reactions.

Researchers have several system parameters at their disposal to influence the outcome of the reaction. These comprise feed concentrations of the reactants, temperature, pressure, as well as a multitude of reactor-specific technical parameters. By recording concentration time profiles under different reaction conditions, information regarding the reaction kinetics can be deduced. In mean-field chemical kinetics the rate of a given reaction is generally expressed in form of rate laws for the time-dependent concentrations $[\mathrm{X_i}](\mathrm{t})$ of the $1 < i < N$ reaction components (reactants and products),

$$\frac{d[\mathrm{X_i}]}{dt} = \sum_j k_j M_{ij} \prod_m ([\mathrm{X_m}]^{\mathrm{s_{jm}}}) \quad , \tag{3.1}$$

where k_j are rate constants, M_{ij} are the elements of the stoichiometric matrix and s_{jm} are the stoichiometric coefficients of the reactants [35, 37]. Even for complex reaction mechanisms, these rate laws can often be asymptotically approximated by effective rate equations with only a few dominating terms. An example would be

$$\frac{d[\mathrm{P}]}{dt} \propto [\mathrm{A}]^{\nu_\mathrm{A}} \cdot [\mathrm{B}]^{\nu_\mathrm{B}} \quad , \tag{3.2}$$

where $[\mathrm{P}]$ refers to the concentration of the product depending on the reactant concentrations $[\mathrm{A}]$ and $[\mathrm{B}]$. The powers ν_A, ν_B are then referred to as (partial) apparent reaction orders. These simplified equations are generally only valid in certain reaction regimes. A famous example for such effective equations is the Lindemann mechanism [90]. It describes the self-activation of gaseous species with a subsequent reaction of the activated species. Under low pressure conditions this reaction can be described by an effective second-order law, while under high pressure the reaction is dominated by the subsequent step leading to an effective first-order dependency.

A variety of methods has been developed in order to deduce such effective kinetic equation from concentration and rate versus time data. Particularly in heterogeneous catalysis, such methods often start off from a detailed microkinetic model [5], either in mean field approximation [39] or within surface arrangement-resolved kinetic Monte Carlo [4]. Establishing such a microkinetic model requires a detailed mechanistic understanding of the reaction though, the lack of which is typically one of the prime motivations for establishing effective kinetic mechanisms in the first place. Furthermore, the more complex the reaction network, the more kinetic parameters like activation barriers or exponential prefactors need to be determined on the basis of experimental data. This can lead to questionable inference on the model parameters or even overfitting due to the lack of available data [91]. In homogeneous catalysis, Blackmond [92] established the concepts of reaction progress kinetic analysis (RPKA) and graphical rate equations in 2005. This analysis investigates the dependence of the reaction rate on regime-determining parameters like reactant concentrations.

Building on these ideas, methods like the variable time normalization analysis (VTNA) by Burés *et al.* [93–95] and nonlinear fitting techniques [82, 96] have been introduced. They come with the advantage of directly operating on the concentration time profiles to obtain effective reaction orders in the individual components. Additionally, such analyses have proven to be robust against sparse and noisy data, straightforward to interpret and simple to perform [93]. There are, however, also some severe caveats. Notably, a separate analysis of the kinetics in each reactant completely neglects any potential interaction between multiple species. In experimental design [2], parameters which influence the outcome of an experiment and can be controlled by the experimenter are referred to as factors. In order to estimate the interactions of these factors, a proper factorial design is required that varies multiple factor settings at the same time. Only if this or any prior knowledge of the system shows that any potential interaction can be excluded, a separate analysis of all factors is valid.

In general, many of the existing kinetic modeling techniques thus suffer from *a priori* model assumptions. The latter are prone to introduce systematic errors and thus a non-normal error distribution in the model predictions, in contrast to the Gaussian distributed error that would be expected from statistical measurement errors. If the model structure does not fit the physical problem underlying the data, the distribution of prediction errors will correspondingly be skewed. This may be due to an important reaction pathway missing in the assumed mechanism, but also unnecessarily complicated models can be a source of error. An analysis of the residual distribution would readily reveal such issues, but unfortunately it is often neglected. In cases where the residuals are indeed provided [82, 91], systematic deficiencies of the models become obvious immediately.

Within this perspective, we here introduce an algorithm for the systematic investigation of formal kinetics from experimental data. Effective rate laws are determined from concentration profiles acquired at different reaction conditions. Performing these measurements according to statistical design [2] allows the algorithm to quantitatively evaluate interactions between different experimental parameters. Centrally, applying well established statistical methods [8, 55, 57] we directly incorporate the goal of normally distributed errors into the optimization. At the same time, the algorithm determines relevant terms in the model function automatically [6, 45], making the approach to a certain degree free of *a priori* model assumptions.

3.1.2 The Modeling Algorithm

The apparent reaction order can be understood as a measure of the sensitivity of the reaction rate to a change in reactant concentration [93]. Similar, sensitivity parameters can also be defined for other reaction conditions and are sometimes interpreted as physical quantities. The algorithm, proposed below, is designed to determine effective rate laws by optimizing these sensitivity parameters. This basic idea is closely related to VTNA. In VTNA, the determination of the order in a certain reactant starts with a set of concentration versus time profiles at different reactant concentrations. Then, if the time axes for all profiles are normalized by the integral over the reactant concentrations raised to the correct power, all profiles should coincide. Looking at eq. (3.2), we can identify this power (or exponent) as the apparent reaction order. It is the sensitivity parameter of the reaction rate with respect to the reactant concentration. In VTNA, the correct value is typically obtained manually in a trial and error process or in a graphical manner, as coinciding curves can be easily identified by visual inspection.

In contrast, our algorithm uses a more quantitative criterion to identify the correct sensitivity parameters. VTNA works, because the relation between the change in product concentration and the reactant concentration becomes linear, if raised to its respective reaction order. How linear a relation between two variables is, can easily be analyzed by performing linear regression and observing the quality of the resulting model. A well established method, based on the same principle, is the Arrhenius plot. There, both temperature and rates are scaled by a nonlinear transformation revealing a linear relationship with a slope equal to the activation barrier of the process. In the case of the Arrhenius problem, the nonlinear transformations are known from the underlying theory and we are interested in the slope of the resulting linear model. In contrast, if we want to determine effective rate laws, it

is known that there should exist a set of transformations, which leads to a linear relationship. Yet, this set of transformations is not easily accessible without detailed knowledge of the reaction mechanism. Notwithstanding, knowing that a linear relationship should exist, the quality of the regression model can be used as a measure of linearity and we can formulate an optimization problem to automatically identify reaction orders and sensitivity parameters for various factors influencing the reaction. These sensitivity parameters are the key quantity in our approach, as they are supposed to capture all nonlinear effects encountered in reaction kinetics.

We assign one of these sensitivity parameters to every factor, which influences the rate of the reaction. For reactant concentrations, these parameters correspond to their respective partial reaction orders. And, as we will show in a later section, also the apparent barrier can be identified as such a sensitivity parameter by rescaling the reaction temperatures. In the regression model we want to set up, these parameters are the exponents, which introduce nonlinear effects to every factor. So, if we vary the concentration of reactant A ([A]) in a reaction, the corresponding quantity, which enters our model, would be

$$\{\mathrm{A}\} = [\mathrm{A}]^{\alpha_\mathrm{A}} \quad , \tag{3.3}$$

where $\{\mathrm{A}\}$ denotes the transformed reactant concentration and α_A is the corresponding sensitivity parameter. As this is a type of power transformation [55], we will also refer to these sensitivity parameters (α) as transformation parameters. If reactant A contributes linearly to the reaction rate, e.g. in a simple first order reaction, the corresponding α would be 1. For a second order reaction in A, on the other hand, the optimal α should be close to 2.

Model function As long as we are only interested in a single influencing factor, for example we only want to vary the concentration of one reactant, a linear model is sufficient to describe the concentration profile, given the correct scaling of the input and output variables. In a more general case, however, we have multiple reactants and additional factors like pressure or temperature which influence the rate of a chemical reaction. Further, we would like to investigate the influence of these multiple factors on the production rate of possibly even multiple product species. Therefore, we need a more complex model ansatz, which allows for possible interactions and higher order terms. For this reason, our algorithm relies on polynomial functions to describe effective kinetic models. Let us consider a simple reaction with two reactants A and B forming the product P. The only factors we want to account for are the concentrations of these two reactants. In that case, the corresponding polynomial would be

$$\frac{d[\mathrm{P}]}{dt} = \text{const.} + k_1 \cdot \{\mathrm{A}\} + \mathrm{k}_2 \cdot \{\mathrm{B}\} + \mathrm{k}_3 \cdot \{\mathrm{A}\} \cdot \{\mathrm{B}\} + \mathrm{k}_4 \cdot \{\mathrm{A}\}^2 + \mathrm{k}_5 \cdot \{\mathrm{B}\}^2 + ... \tag{3.4}$$

with k_i the unknown coefficients of the respective polynomial terms, which need to be determined in the regression. Note that this is now a function of the transformed factors, meaning the transformation parameters (α) are already contained in the terms of this sum. Such a polynomial contains not only higher powers of the input variables but, most importantly, also the necessary interaction terms between multiple factors, meaning e.g. the product $\{\mathrm{A}\} \cdot \{\mathrm{B}\}$. At the same time, polynomial regression is still linear in the unknown parameters, meaning that the quality of the model can still be used as a measure of adequacy for the determined sensitivity parameters. However, being too rigid with the functional form of our model may introduce systematic errors. To account for this we use the least absolute shrinkage sparsification operator (LASSO) [6, 45], to let the the algorithm itself decide, which polynomial term to include in the final model. LASSO is a modification to the standard linear regression, which, by introducing a regularization term, is able to set regression coefficients to 0. This way, all terms which do not significantly contribute to the solution based on the available data, are automatically canceled out. Further, the remaining (non zero) coefficients not only tell us, which polynomial terms contribute to the effective model. The magnitude of the LASSO coefficients give us a measure of the relative importance of different polynomial terms or, in this context, kinetic processes. Additionally, the sign of the coefficient indicates a positive or negative influence on the reaction rate.

Despite selecting only the significant terms from the polynomial ansatz, we do not want to expand eq. (3.4) to arbitrary order, as this would result in a huge amount of available terms to choose from. The final result will depend on the order up to which these ansatz terms are expanded and not every expansion order is reasonable for every problem. Hence, we need to define some guidelines to choose an appropriate ansatz order. First of all, the order of the ansatz terms should generally not exceed the number of factors in our reaction. If we only consider two reactant concentrations for our model, an expansion up to 3rd order would be pointless, as we can not define any 3-factor interactions in that case. Including e.g. temperature as a third factor, such 3rd order terms could become relevant and should be included in the ansatz. Further, we only consider interactions (or multilinear) terms, meaning product terms of multiple different factors like $\{\mathrm{A}\} \cdot \{\mathrm{B}\}$ as opposed to the purely quadratic ($\{\mathrm{A}\}^2$) or higher power ($\{\mathrm{A}\}^3, \{\mathrm{A}\}^4, ...$) terms. The reasoning behind this is that in some cases these terms are already captured by the power transformation in the original factors. For example, a quadratic term with transformation parameter $\alpha = 1$ is equivalent to a linear term with $\alpha = 2$

$$\left([\mathrm{A}]^1\right)^2 = \{\mathrm{A}\}^2_{\alpha_\mathrm{A}=1} = \{\mathrm{A}\}_{\alpha_\mathrm{A}=2} = \left([\mathrm{A}]^2\right)^1 \quad . \tag{3.5}$$

This, of course, does only hold for very simple systems with only one polynomial term containing [A]. There are exceptions, where including the purely quadratic terms can be beneficial though, for example, when the target property (here the reaction rate) shows an extreme point within the region, we want to describe and we will further illustrate and discuss this point below. In general, however, for an initial analysis we neglect these terms. We focus on the remaining pair and higher-order interaction terms. These are especially relevant for factors, which are reactant concentrations. In classical collision theory [97], the collision probability of multiple species directly depends on the product of their respective concentrations. This directly transfers into chemical kinetics as a collision (or at least spatial proximity) is an essential requirement for any kind of reaction to take place. This is the reason, why these interaction terms often appear in kinetic rate laws and hence we need to consider them in our model ansatz. The probability of many-body collisions, of course, depends on the type of reaction at hand. Orders of four or higher, however, do usually not contribute significantly, and thus provide a natural upper bound. Another important point is the amount and type of data available, limiting the number of parameters, which can reasonably be estimated. Adding higher order terms also introduces additional degrees of freedom, giving the algorithm more flexibility. This can result in a more accurate model, which in turn will be less interpretable due to a multitude of selected terms. There are, thus, various things to consider when choosing the order of the ansatz polynomial. In practice, a reasonable rule of thumb is to choose a 2nd order ansatz for two input features and a 3rd order ansatz for more than two features. A schematic, showing how the initial factors translate into the target property for the simple example $\mathrm{A} + \mathrm{B} \rightarrow \mathrm{P}$, is given in Figure 3.2.

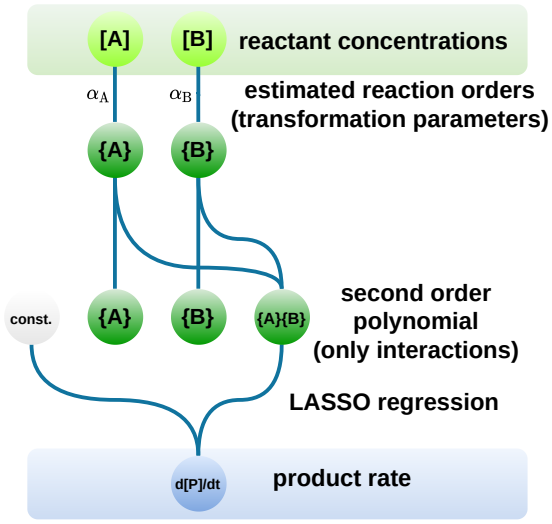


Figure 3.2: Model structure for the relation between reaction factors and product rate. The factors (or input features), in this case, are the concentrations of the reactants. This is an example for a reaction with two reactants A and B and a single product P. The reactant concentrations are transformed by the current guess of their respective reaction orders (transformation parameters α). Based on these transformed variables, a polynomial model is constructed. In the regression, LASSO selects only those terms, which significantly contribute to the target variable, here the product rate.

Cost function For the optimization of the sensitivity parameters, a quantitative objective is necessary. We want to obtain a statistically robust model, which at the same time achieves a reasonable fit of the provided data. The quality of the fit could be quantified by simply looking at the residual norm of the model. However, at the same time, we want to make sure, that the LASSO regression does not introduce systematic errors into the model by choosing unreasonably complex polynomial terms to enforce small residuals. To achieve this, we explicitly consider the distribution of the model residuals in our cost function. As we assume the statistical noise in the underlying data to be normally distributed, we also want the residuals of our model to show such a distribution. A normally distributed model error implies that everything the model can not explain is to be assumed as statistical noise. To quantify this criterion, the probability plot correlation coefficient (PPCC) introduced by Filliben [7] is used (cf. section 2.5). The PPCC quantifies the normality of a statistical distribution as a value between 0 and 1. The closer to 1, the closer the distribution is to being normal. We then end up with two separate objectives for the optimization, the quality of the fit and the normality of the error distribution. This corresponds to a vector valued cost function which can be written as

$$\min_{\boldsymbol{\alpha}} \left\{ |R|_{Fro}(\boldsymbol{\alpha}), -PPCC(\boldsymbol{\alpha}) \right\}, \tag{3.6}$$

where $|R|_{Fro}$ is the Frobenius norm of the residual matrix, which for a one dimensional residual vector is equal to its euclidian norm. The negative sign in front of the PPCC denotes, that we want to maximize the PPCC value to get as close to normality as possible.

Estimation of the transformation parameters A common approach to solve a multi objective problem like equation 3.6 is to introduce a scoring function, which translates the objective vector into a scalar. For example, we could minimize the norm of the objective vector. Such a scalar score then allows for the use of standard optimization techniques. A severe drawback, however, is that we introduce some bias by weighing the different objectives in the scoring function. To avoid introducing such a bias in this step of the algorithm, we utilize the multi-objective evolutionary algorithm NSGA-II [8]. A multi-objective optimization usually does not result in one optimal set of parameters satisfying all objectives at the same time. Instead, we rather end up with a set of possible models forming the so called pareto front (PF) as illustrated in Figure 3.3. These pareto-optimal models display different

trade offs between the multiple objectives, where an increase in one objective would lead to a lower score in another. Hence, it is not straightforward to consider one pareto-optimal model to be better than any other one, as it is not possible to find a model with a higher score in all objectives.

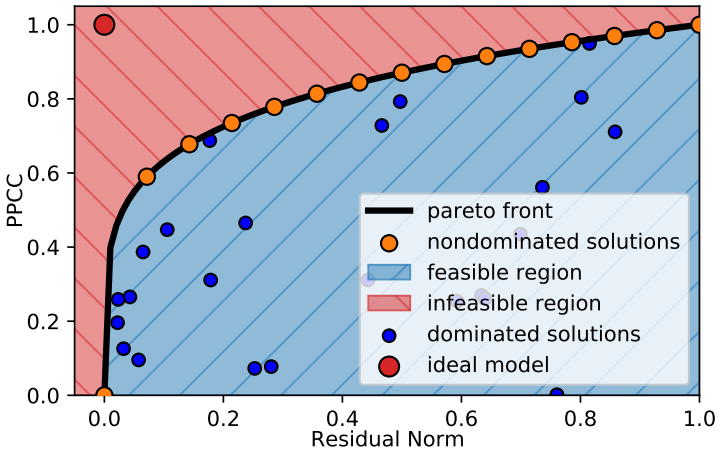


Figure 3.3: Schematic of the pareto front for a kinetic modeling problem. The two objectives are the PPCC value of the model residuals and the norm of the residual vector.

In this context, the perfect model would achieve a PPCC value of 1 and a residual norm of 0. Looking at the schematic pareto front in Figure 3.3, we see that these two objectives interfere with each other. The ideal model would be located in the upper left corner of the plot, within the infeasible region. Points on the pareto front, close to this ideal model, would provide a reasonable choice. Solutions with even lower residuals can only achieve this at the cost of neglecting the error distribution. Going to the other end of the pareto front, there are models which achieve an almost perfect normal distribution of errors but can not provide a reasonable fit anymore. Ultimately, the solution offering the best trade off between all objectives needs to be chosen based on additional criteria and information. Below, we will discuss characteristic features of the pareto front, that can be taken as strong indicators for valid choices. A schematic, summarising all relevant steps of our approach, is shown in Figure 3.4.

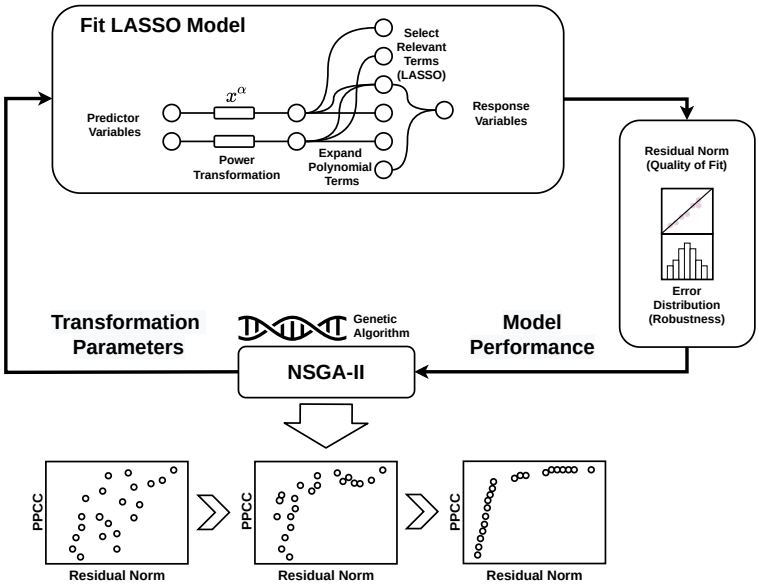


Figure 3.4: A schematic of the optimization algorithm for obtaining the model transformation parameters. Using the NSGA-II algorithm, we optimize the model parameters for the power transformation, which enter the polynomial model and the LASSO regression. For each LASSO model, both the residual norm and distribution of errors (in terms of the PPCC value) are determined and used as the cost function during the optimization. While the initial parameter candidates are chosen randomly, during the optimization the pareto optimal solution emerge.

Pareto Front Analysis and Model Selection The huge amount of pareto fronts generated in the temperature scan analysis mandates an automatic model selection process. In this work a rudimentary method was used. Basically, we analyze the distances between neighboring pareto optimal points along the residual axis. We determine the pairwise distances of all subsequent points and select those pairs, which are separated by more than k (default $k = 3$) times the average distance. This way we locate the gaps in the front. From these gaps, we choose the one at the lowest residual value. Here we select the final model again to be the data point on the smaller residual end of this gap. In case the pareto front is approximately continuous with no significant gaps along the residual axis, the algorithm selects the model closest to the median of the residual values of all solutions.

3.1.3 Implementation Details

Software packages The entire optimization algorithm is implemented in Python. It relies on the NSGA-II implementation as provided in the Pymoo [98] multiobjective optimization package. The implementation of the LASSO algorithm from scikit-learn [99] is used and the PPCC calculations are based on scipy [100].

Data Scaling The optimization of the transformation parameters is not directly performed on the raw data provided as input for the algorithm. Before the power transformation is applied, the data (both predictor and response data) is shifted to all positive values. This is done, as the Tukey power transformation is not defined for negative data. The same holds for the Box-Cox transformation [2]. The algorithm also has the option to use the Yeo-Johnson transformation [58], which can deal with negative data. Further, when dealing with reaction kinetics, concentrations of the reactants as well as product rates should in general have positive values. For the sake of generality, however, the shift of the data is always performed. Between the power transformation and the regression task, all individual predictors are scaled to zero mean and unit variance to eliminate any bias in the LASSO regression caused by the absolute quantity of the feature. The standardization of these predictors leads to the LASSO coefficients mainly having values in the interval [-1,1]. All these steps are linear transformations on the data, which can be reverted as along as the shift and scale parameters are

stored. The initial shift of the data to positive values has also implications on the extrapolation range of the fitted models. During the procedure, the minimal data value for every feature is shifted to a predefined minimal value (default zeroshift = 1). The applied shift is determined based on the training data. When we predict for feature values outside the training range, the applied shift may lead to negative values. In case of some transformation parameter (α) smaller 0, the resulting model prediction will be undefined. This problem, in principle, can be eluded by increasing the zeroshift value. Unfortunately, shifting the data to larger absolute values decreases its sensitivity to the transformation parameters leading to a worse fit. An example is given for the CO oxidation model in the appendix (cf. Figure A.2).

LASSO Regularization Strength The LASSO regression requires a regularization strength parameter as input. Usually, this parameter is determined using e.g. a k-fold cross validation. To avoid repeating this cross validation in every iteration for every individuum in the NSGA-II optimization (which would drastically increase the computational cost), we decided to determine a global regularization strength once in the beginning with all transformation parameters (α) equal to 1. This is then kept fix for the entire optimization. This way, we also avoid possible oscillations with parameters jumping back and forth in every iteration, hindering the convergence of the genetic algorithm.

Temperature Input When we want to use the algorithm to identify apparent activation barriers by varying the temperature we make use of the following relation:

$$k = k_0 \cdot \exp \frac{-E_A}{R \cdot T} = k_0 \cdot \left(\exp \frac{-1}{R \cdot T} \right)^{E_A} = k_0 \cdot (\exp -\beta_M)^{E_A}, \beta_M = \frac{1}{RT} \tag{3.7}$$

This allows us to identify the apparent barrier as the sensitivity parameter of the $(\exp -\beta_M)$ term. As these sensitivity parameters are exponents applied in a power transformation, we have to make sure, that their numerical values lie within reasonable boundaries. For this reason, we need to scale the universal gas constant R to a suitable energy unit, such that the expected exponent does not become too small or too large. For example, if the expected barrier lies in the range of 1 kcal/mol to 20 kcal/mol, we would insert R in units of $\frac{10kcal}{mol \cdot K}$. The exponent determined in the algorithm then should have a value between 0.1 and 2. When rescaling R to a suitable unit, we often end up with extremely small values for $(\exp -\beta_M)$ (e.g. on the order of 10^{-32}). Such values can be rescaled by a constant to ensure numerical stability. Such a scaling should not change the relative curvature between predictor and response and hence not affect the determined activation barriers.

$$[c \cdot \exp(-\beta_M \cdot q)]^{\frac{E_A}{q}} = c^{\frac{E_A}{q}} \cdot [\exp(-\beta_M \cdot q)]^{\frac{E_A}{q}} \tag{3.8}$$

where q and c are constants scaling the feature values to a reasonable numerical range without changing the curvature, as the prefactor does not depend on β_M. By choosing q in terms of an energy conversion factor the fraction $\frac{E_A}{q}$ becomes dimensionless. Another side effect of the exponential sensitivity of the rates on temperature is that the scaled temperature values $(\exp -\beta_M)$ may span several orders of magnitude, if the temperature region of interest is large. For example, a temperature range of 200 K in some cases can lead to a span of 8 orders of magnitude in $(\exp -\beta_M)$. This huge range again can lead to numerical instabilities during the optimization. On the other hand, small temperature ranges may become an issue from an experimental point of view, as the accuracy of thermostats is limited, especially at elevated temperatures. We found a range of around 100 K to be a reasonable trade off between these two boundaries. ´

Power Transformation of the Response In some cases, nonlinearities on the predictor variables are not enough and the functional relationship can be simplified when also considering a transformation of the response variables. Again, we can apply the Tukey-transformation and obtain:

$$\frac{d[\mathrm{P}]}{dt} = (\text{const.} + k_1 \cdot \{\mathrm{A}\} + k_2 \cdot \{\mathrm{B}\} + k_3 \cdot \{\mathrm{A}\} \cdot \{\mathrm{B}\} + ...)^{(\beta)} \tag{3.9}$$

We assume the transformations on the response (β) to be robust over all possible model functions, meaning we only have one global set of response transformations for the entire population during the NSGA-II optimization. Therefore, we do not explicitly optimize β, but instead estimate every few iterations, which β is on average the best possible transformation over the entire population. This way, we not only reduce the dimensionality of the problem, but also end up with a more stable optimization, as we do not play off two nonlinear transformations against each other.

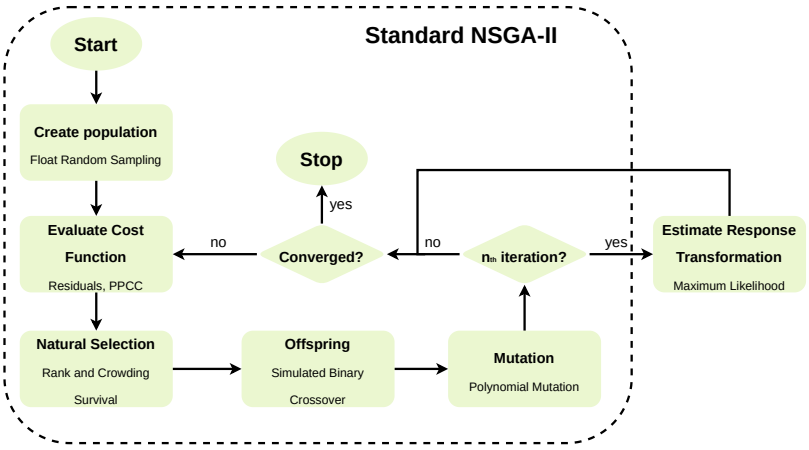


Figure 3.5: Schematic of the NSGA-II optimization including the slight modification for estimating the transformations of the response variables.

While such transformations of the response are implemented in the modeling algorithm, the results presented in this chapter do not require this additional nonlinearity. However, in the following chapters, approximate models will include this response transformation.

3.1.4 Validation Test Case: Classical Examples of Formal Reaction Kinetics

Generation of Artificial Data Sets Synthetic concentration versus time data is obtained by numerically integrating the partial differential equations given by the reaction rate laws for these processes. A simplified one-dimensional transient plug flow reactor (TPFR) model is simulated, applying the method of lines [101] for the discretization along the reactor axis. Numeric approximations to the time derivatives of the concentrations have been obtained from central differences. For this purpose, the reaction has been simulated at different flow rates. In our idealized reactor model, a change in flow rate is equivalent to a change in residence time. At constant volume (reactor length), a change in residence time in a TPFR model is in turn equivalent to a change in reaction time (time before quenching) in a batch reactor. Therefore, in the limit of short residence times, we can estimate the reaction rate as the change of the product concentration at the reactor exit.

$$\begin{aligned} \frac{d[\mathrm{P}]}{dt} &\propto \frac{d[\mathrm{P}]_{exit}}{d\tau} \\ \frac{d[\mathrm{P}]_{exit}}{d\tau} &\approx \frac{[\mathrm{P}]_{exit}(\tau+\delta\tau) - [\mathrm{P}]_{exit}(\tau-\delta\tau)}{2\delta\tau} \end{aligned} \tag{3.10}$$

where t denotes the simulation time and τ refers to the residence time in the reactor. $\delta\tau$ is the finite change in residence time used to approximate the derivative. As we are not interested in the absolute values of the rates, in this work these derivatives with respect to the residence time τ will be used as approximations to the actual rate. The examples discussed in the following sections are all evaluated for short residence times and thus describe the low conversion limit of the respective chemical processes. In this limit, the overall concentrations are still very sensitive to the chemical reaction kinetics rather than to the transport through the reactor tube. After all, the effective model we construct are designed to only describe the

reaction kinetic part of the reactor dynamics. Consequently, we generate training data for our algorithm in a way, that is most sensititive to kinetic quantities.

The proposed algorithm is designed to operate on experimentally measured data. According to the central limit theorem [102], the accumulated effect of various noise sources in an experimental setup tends to result in a normally distributed measurement error. This gets more and more pronounced, the larger the data set is. As our algorithm relies on such normally distributed noise, we need to ensure that this distributional assumption is also valid for our artificial data sets. To simulate statistical noise of the experimental setup, a noise term based on normally distributed random numbers is therefore added to the computed rate data. The magnitude of this noise term is 1% of the mean value of the respective rate data set. Measurement errors for reaction rates around 1% to 2% are encountered in literature [103, 104] and a value of 1% lies well within the region, where the fitted rate models are not significantly affected by the noise. Details on the influence of the magnitued of the noise are given in the appendix.

The resulting data sets consist of the reactant concentrations at the reactor exit as well as the change in concentration or rate of the product species at different reaction conditions. In the examples shown in this work, these reaction conditions are characterized by differences in the feed composition (reactant concentrations) and temperature. It is important to properly sample these conditions, in order to be able to capture especially the interaction effects at play. For this reason we apply factorial experimental designs [2]. In these factorial designs, instead of keeping all but one factor fixed and vary each individually, all factors are varied at the same time. The simplest example of such a design is called a full factorial design (FFD). In a FFD, all possible combinations of factor settings are sampled. Again, we look at the example of the simple reaction $A + B \rightarrow P$. If we want to vary the feed concentration of A and B, we first have to decide, how many settings we want to allow. Suppose we want to look at three different concentrations of both A and B, the corresponding FFD would dictate $3 \cdot 3 = 9$ necessary experiments to sample all possible combinations. However, the amount of experiments for a FFD quickly becomes infeasible, if we increase the number of levels or factors. Therefore, such FFDs are only applicable for simple experimental setups or, in our case, simple reaction examples. The advantage of the FFD is that it contains the maximum amount of information possible about higher order and interaction effects. For more complex problems, we can often reduce the number of experiments by sacrificing information on effects beyond a certain order. Such designs are then generally referred to as fractional factorial designs. Most of the test cases in this study deal with simple reactions with only two to three factors. For these cases, we will use FFDs to sample the respective reaction conditions. In a later example, we will make use of a very efficient screening design, to reduce the number of required experiment for a slightly more complex chemical system.

The Isothermal HBr Reaction As a first test, we apply our algorithm to determine an effective rate law for the hydrogen bromine reaction at constant temperature. A set of concentration and rate data obtained at 25 different feed compositions is used to deduce the reaction orders in both reactants. This data set as well as the underlying data sets for the following sections are explicitly provided in the SI.

As we deal with two factors (the concentrations of H_2 and Br_2) in this first example, we choose a 2nd order ansatz for the isothermal HBr problem. Feeding reactant concentrations and product rates into our algorithm, we obtain the pareto front shown in Figure 3.6 a).

Table 3.1: Model parameters and objective values for five selected pareto-optimal HBr models from the pareto front in Figure 3.6.

	$\alpha_{[Br_2]}$	$\alpha_{[H_2]}$	Residual	PPCC	dominating LASSO term
▶ Solution 1	0.658	1.273	0.259	0.961	$\{Br_2\} \cdot \{H_2\}$
▶ Solution 2	0.518	0.985	0.321	0.989	$\{Br_2\} \cdot \{H_2\}$
▶ Solution 3	0.486	0.916	0.352	0.995	$\{Br_2\} \cdot \{H_2\}$
▶ Solution 4	-4.066	2.908	1.258	0.995	$\{H_2\}$
▶ Solution 5	-7.121	4.511	1.867	0.996	$\{H_2\}$

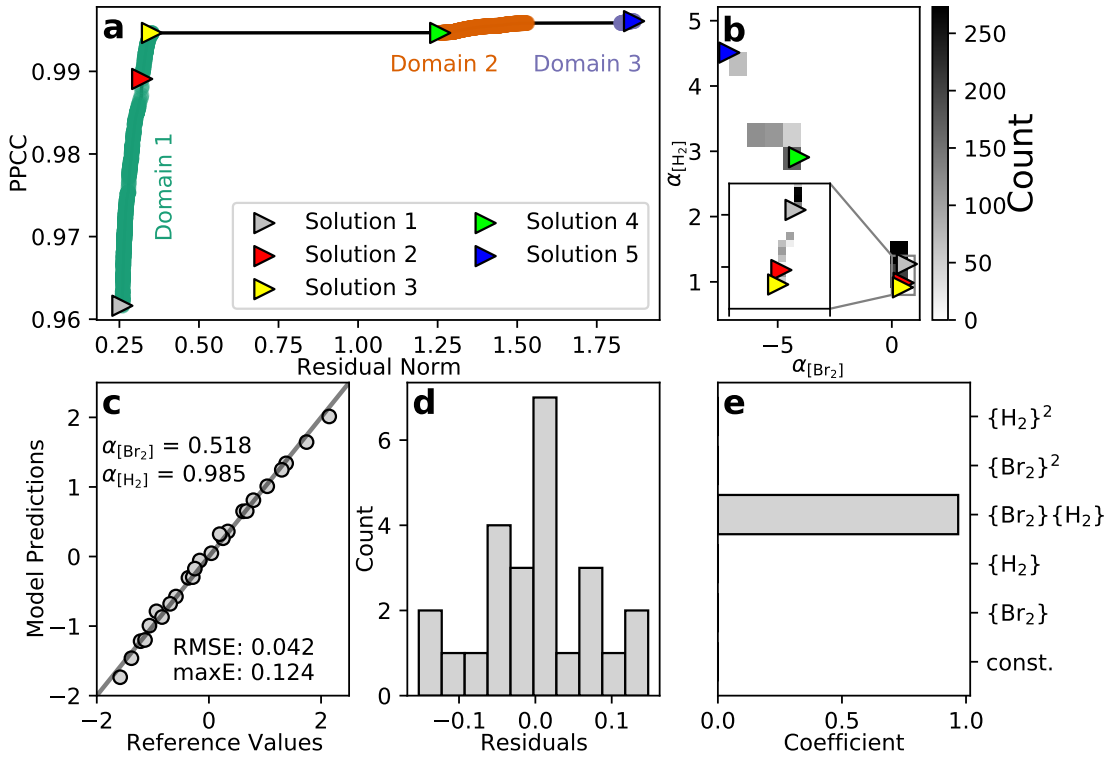


Figure 3.6: a) Pareto front for the isothermal HBr problem. The NSGA-II optimization is performed on a population of 1000 individuals. Five representative pareto-optimal solutions are highlighted with their respective effective reaction orders and model parameters given in Table 3.1. b) Distribution of reaction orders α for both H_2 and Br_2 over the entire population. c) Parity plot (predicted reaction rates vs. reference rates from microkinetic model) for solution 2. d) Residual distribution for solution 2. e) LASSO regression coefficients for solution 2.

Each point on the pareto front resembles a possible solution to the modeling problem. These solutions differ in the transformation parameters α on the reactant concentrations, i.e. the apparent partial reaction orders, and they differ in the set of dominant model terms selected by LASSO. Recall that these model terms are combinations of reactant concentrations which themselves are subject to a nonlinear transformation through the reaction order. For the isothermal HBr problem, the pareto front is composed of three distinct domains of solutions. The majority of solutions show comparably small residuals and are spread over a wide range of PPCC values. The other two domains show a slight increase in PPCC at the cost of a significant increase in model residuals. The transformation parameters, i.e. the reaction orders in the dominating regression term, for five representative solutions are given

in Table 3.1. Solutions 1, 2 and 3 are located within the major domain of the front. All three models show reaction orders for Br_2 close to 0.5 and for H_2 roughly equal to 1. They thus reproduce the effective rate law from Bodenstein and Lind [78] with the underlying radical reaction mechanism. Solutions 4 and 5 are located in other domains of the front. They show vastly different parameter sets which cannot be reasonably interpreted as reaction orders.

Figure 3.6 b) shows that this structure of the pareto front is also resembled in parameter space, where the distribution of possible reaction orders is arranged in three domains. This means, that all solutions within one domain of the pareto front show very similar transformation parameters. Switching from one domain to another does not only go with a change in parameters. Also the dominating LASSO coefficients change, meaning the model has an overall different functional form. For all solutions located in domain 1, only the interaction term $\{Br_2\} \cdot \{H_2\}$ contributes predominantly to the rate law, while the largest coefficient in solutions 4 and 5 corresponds only to the concentration of hydrogen $\{H_2\}$. This behavior exemplifies the flexibility of our approach. As soon as all possible trade-offs for a fixed functional form are exhausted, LASSO allows for the switching to an alternative model which can provide additional pareto-optimal solutions. This discontinuous structure of the pareto front also allows for a selection of interesting solutions without prior knowledge on the desired result. Pareto solutions located close to a discontinuity of the front resemble an extremum. These points describe either the maximum PPCC value, or the minimal residual norm for a given set of LASSO terms. Changing the transformation parameters (α) any further, would result in the LASSO optimization converging towards another set of terms and hence a different functional form of the resulting rate law. Further, the ideal solution should show a high PPCC value while maintaining a reasonably small residual norm. For most pareto fronts generated by our algorithm, these solutions are located in the upper left corner (e.g. in Figure 3.6). For the HBr test case, the pareto front shows a clear kink in this region, followed by a large discontinuity. Following this reasoning for a chemically reasonable region of the pareto front, we should find the best possible solution at the upper edge of domain 1. And indeed solution 3 does show the expected reaction order of almost exactly 3/2. In the following sections we will show, that following such a heuristic does also result in chemically reasonable solutions for other systems, where we do not know the expected outcome in advance.

A closer look at the prediction quality of solution 2 is shown in Figure 3.6 c). The partially negative values for the reaction rate in this figure result from the rescaling and shifting of the data during the modeling procedure. This is a technical requirement and details are given in the SI. Within the limits of 25 available data points, the error distribution for this model shown in Figure 3.6 d) roughly resembles a Gaussian shape, resulting in its high PPCC value. While the LASSO regression is relatively robust against small sample sizes, evaluating this PPCC value will become the bottleneck of the proposed algorithm with respect to the required amount of data points. As the PPCC value ultimately compares two distributions, its significance depends on how well the shape of these distributions is captured by the available data. While the quality of the PPCC increases with larger data sets, it is generally capable of distinguishing between different distributions with remarkably small data, way below 100 points [7]. The 25 data points used in this example, however, are probably close to a minimum size in order to obtain reasonable results. A closer look at the convergence of the PPCC value with respect to the number of data points is provided in the SI. Further, the model error is small with a relative root mean squared error (RMSE) of 4.2%. Another interesting result is, that the LASSO algorithm chooses exactly only one polynomial term for this model of the HBr rate, perfectly reproducing the literature known effective kinetic law.

Estimating Temperature Effects in the HBr Reaction In a second data set, additionally also the temperature of the Hydrogen Bromine reaction is varied to illustrate the effect of temperature as another factor. A total of 45 concentration profiles were full factorially simulated within a temperature range between 840 K and 860 K. This temperature window is chosen to be that small, as the HBr reaction shows a very high sensitivity with respect to changes in temperature. Keeping the temperature variations small results in an effect on the reaction rate of similar magnitude compared to variations in feed concentrations. This prevents one factor from dominating the LASSO regression. Nevertheless, larger temperature ranges can also be realized within certain limits. A detailed discussion on this, as well as

the complete data set are provided in the SI. We can, however, not directly use the temperature data in order to determine its effect in our approach. If we assume an Arrhenius-like behavior for the temperature dependency of a general rate constant k, we can rearrange the corresponding Arrhenius equation as follows

$$k = k_0 \cdot \exp\left(\frac{-E_A}{R \cdot T}\right) = k_0 \cdot \left(\exp\left(\frac{-1}{R \cdot T}\right)\right)^{E_A} = k_0 \cdot \left(\exp\left(-\beta_M\right)\right)^{E_A} = k_0 \cdot \hat{T}^{E_A}$$
$$\beta_M = \frac{1}{RT}, \quad \hat{T} = \exp\left(-\beta_M\right) \tag{3.11}$$

with k_0 the pre-exponential factor and R the universal gas constant and T the temperature. We define $\hat{T}$ as the rescaled temperature values $\exp(-\beta_M)$ and will use this notation in the following for brevity. By this rearrangement we can identify the apparent barrier E_A as the sensitivity parameter of the reaction rate to a change in $\hat{T}$. Thus, by feeding the algorithm these scaled temperatures additionally to the reactant concentrations, an estimate for the reaction barrier can be obtained. As we now have an additional input feature in terms of the $\hat{T}$ values, we choose a 3rd order polynomial ansatz for this temperature-dependent problem. Figure 3.7 shows the resulting pareto front for this data set. The general structure of this front is similar to that of the isothermal problem.

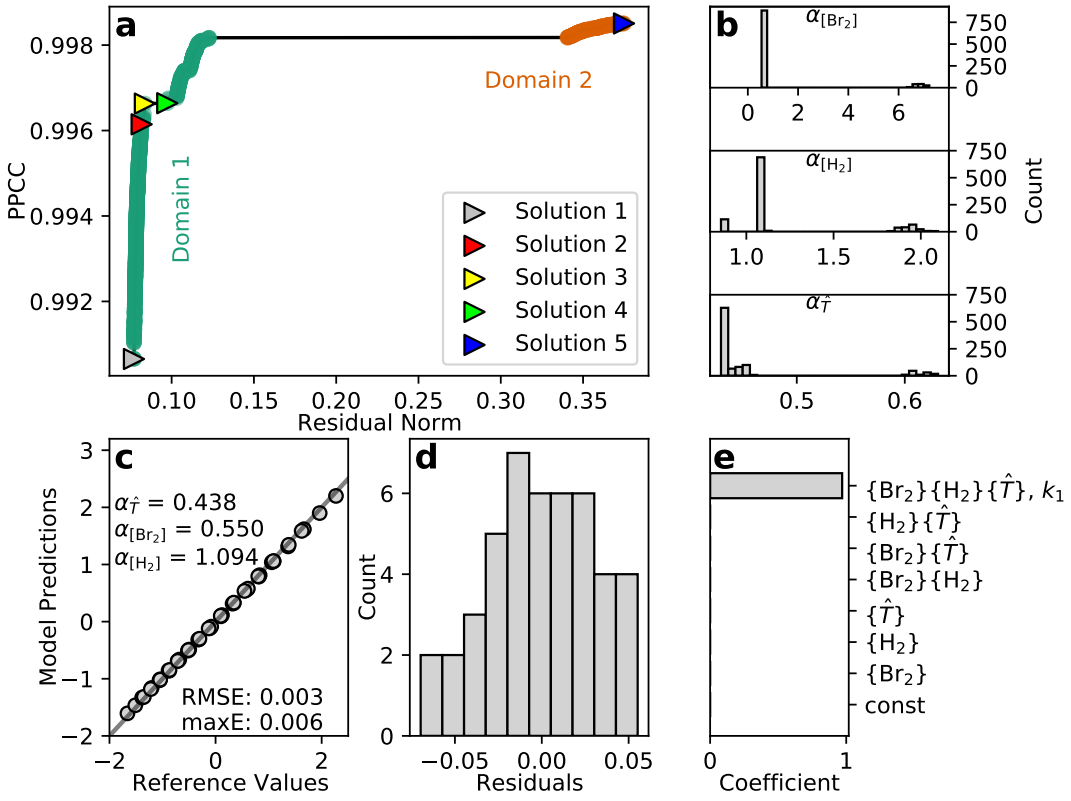


Figure 3.7: a) Pareto front for the temperature-dependent HBr problem. The NSGA-II optimization is performed on a population of 1000 individuals. Five representative pareto-optimal solutions are highlighted with their respective sensitivity parameters given in Table 3.2. b) Distribution of reaction orders α for H_2, Br_2 and the scaled temperature values $\hat{T}$ over the entire population. c) Parity plot for solution 2. d) Residual distribution for solution 2. e) LASSO regression coefficients for solution 2. k_1 indicates the regression coefficient, which significantly contributes to the solution. The resulting approximate rate law is given in eq. (3.12).

Model parameters for five representative solutions are again given in Table 3.2. By means of the heuristics discussed for the isothermal case, solution 2 is located in the chemically reasonable area of the pareto front. Its reaction orders for hydrogen and bromine again

Table 3.2: Model parameters and objective values for five selected temperature-dependent pareto optimal HBr models corresponding to the pareto front in Figure 3.7.

	$\alpha_{[Br_2]}$	$\alpha_{[H_2]}$	$\alpha_{\hat{T}}$	Residual	PPCC
▶ Solution 1	0.546	1.085	0.430	0.077	0.991
▶ Solution 2	0.550	1.094	0.438	0.082	0.996
▶ Solution 3	0.551	1.103	0.442	0.084	0.997
▶ Solution 4	0.546	1.115	0.449	0.097	0.997
▶ Solution 5	7.221	0.860	0.631	0.375	0.999

reproduce the expected literature rate law of 3/2. The apparent activation barrier results in a sensitivity parameter of 0.438. Taking into account the proper rescaling of the $\hat{T}$ values, this inherently dimensionless sensitivity parameter directly translates into an activation barrier of 43.8 kcal/mol. Details on this scaling procedure are provided in the SI.

Levy [80] performed an experimental study on the high temperature kinetics of the HBr formation in the range between 600 K and 1400 K, extending the original study by Bodenstein and Lind [78] which was performed at around 500 K. The artificial data set created in this work was sampled around 800 K well within this high temperature regime. In experiment an apparent activation barrier of around 40.6 kcal/mol [80] was found, which is in reasonable agreement with the 43.8 kcal/mol determined by our algorithm. Looking at the resulting LASSO coefficients for this solution (Figure 3.7 e)), now the 3-body interaction of all factors is dominating. For this solution the effective rate law would correspondingly have the form

$$\frac{d[\mathrm{HBr}]}{dt} \propto k_1 \cdot \left([\mathrm{Br_2}]^{0.550} \cdot [\mathrm{H_2}]^{1.094} \cdot (\exp\left(-\beta_M\right))^{0.438} \right) \tag{3.12}$$

with k_1 the LASSO coefficient for the dominant term as indicated in Figure 3.7 e).

A Homogeneously Catalized Reaction Cycle As a second prototypical case for reaction kinetics we now turn to a simple model catalytic cycle. One or more reactant molecules form an active complex with the catalyst in a pre-equilibrium. This active complex then irreversibly reacts to form the product in a rate-determining step, regenerating the catalyst. To test, whether our algorithm is able to capture the kinetics of such a catalytic process, we create test data based on the model reaction network given in Figure 2.11. The first test deals with the catalytic cycle without any side reaction ($k_6 = 0$). Concentration profiles are obtained for a range of different feed compositions starting at a 1:1 ratio of A to B up to a tenfold excess in B. For each feed ratio, 25 data points based on variations in the feed concentrations of A and B are simulated and an analysis using the modeling algorithm with a second-order ansatz for the considered LASSO terms is performed. For each analysis, one model is selected from the pareto front with the same heuristics regarding discontinuities in mind as discussed before for the HBr problem. The resulting reactions orders in A and B for these selected solutions are shown in Figure 3.8.

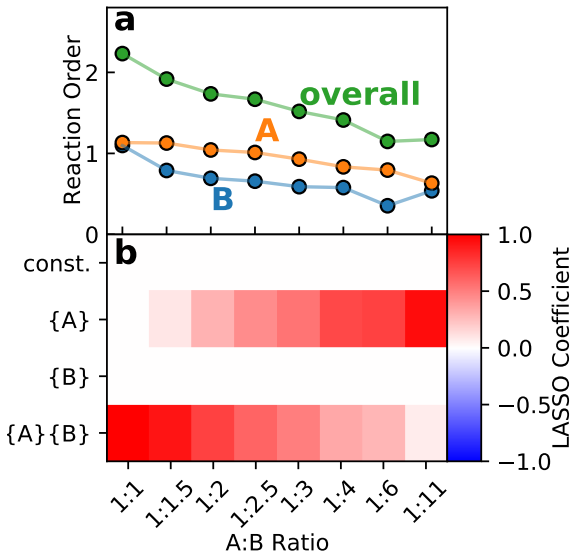


Figure 3.8: a) Effective reaction orders versus excess in reactant B for the model catalytic cycle without side reaction. b) LASSO coefficients for the selected solutions at different A:B ratios. The coloring indicates the magnitude of the coefficients, which corresponds to the significance of this term to the overall solution. At a feed ratio of 1:1 only the interaction between A and B contributes to the corresponding rate law, while with increasing excess in B the linear term in A becomes more and more dominant.

While initially the orders in both reactants are close to 1 at a feed ratio of 1:1, the sensitivity of the reaction rate to the concentration of B quickly decreases to zero. The overall reaction order thus decreases from an initial second-order process to a first order for large excess values. This is to be expected, as ultimately the reaction rate only depends on the concentration of the limiting reactant. This simple example illustrates that the algorithm is capable of identifying such changes in the reaction regime, as clearly visible by the change in reaction orders. Further, as can be seen from Figure 3.8 a), the sensitivities with respect to the individual reaction orders are observed individually. While the apparent order in reactant B steadily decreases with an increasing excess in B, the order in A only shows a slight decrease. The overall reaction order concomitantly switches from initially a second-order process depending on A and B to an effective first-order reaction limited only by A. This way it may easily be deduced which reactant is limiting the overall rate without any initial assumptions. Further insight can be provided by the LASSO regression coefficients, as illustrated in Figure 3.8 b). With increasing excess, the dominating LASSO term switches from the product of both reactant concentrations to purely the concentration of A. This also explains the increasing deviations of the reaction order in B from the expected behavior, as the overall kinetic model becomes less and less sensitive to the concentration of reactant B. In the extreme case of an A:B ratio of 1:11, the term $\{B\}$ does almost not contribute to the overall model anymore and hence also the reaction order in B is not well defined.

Reaction Cycle with Substrate Decomposition In a next step, we introduce the decomposition of reactant A as a side reaction. This leaves us with two target variables, the reaction rate towards product P and side product D. First, we set up a data set of 25 points varying only the feed concentrations at a constant temperature of 700 K. For the model ansatz again 2nd order terms are used, as we have two input features in the concentrations of A and B. It is straightforward to generalize our algorithm to identify surrogate rate laws for multiple targets. The main difference is an increase in the dimensionality of the objective space for the optimization, as now PPCC values for both $\frac{d[\mathrm{P}]}{dt}$ and $\frac{d[\mathrm{D}]}{dt}$ need to be considered. Consequently, this also entails a higher-dimensional pareto front, complicating the selection of a final solution. The resulting pareto front is shown in the SI and exhibits a similar structure as the previously analyzed 2D examples. Hence, at least for this 3D example, a manual model selection by visual inspection was still possible as before.

Table 3.3: Model parameters and objective values for selected models for the production rate of products P and D in the catalytic cycle with side reaction. The corresponding pareto fronts are given in the SI.

	$\alpha_{[A]}$	$\alpha_{[B]}$	$\alpha_{\hat{T}}$	Residual	PPCC
$\frac{d[P]}{dt}$	1.143	1.159	0.066*	0.159	0.995
$\frac{d[D]}{dt}$	1.131	-0.827**	0.051*	0.148	0.981

* in 100 kcal/mol, **not selected in LASSO solution

The optimum solution according to the heuristics discussed before yields the following rate laws for equal feed of A and B

$$\begin{aligned} \frac{d[P]}{dt} &\propto [A]^{1.03} \cdot [B]^{1.07} \\ \frac{d[D]}{dt} &\propto [A]^{1.03} \quad . \end{aligned} \tag{3.13}$$

Again, the algorithm accurately captures the expected behavior. The rate towards the reaction product P is identified to be dependent on both reactants A and B, while only A contributes to the formation of side product D. Additionally introducing temperature effects as done for the HBr example leads instead to complications. In principle, it would also be possible to obtain the two activation barriers for the individual reactions within one model by using the scaled temperatures as an input variable twice. However, to make sure, that only one temperature term contributes to each rate expression, such that the two barriers are disentangled, would result in a constrained LASSO [105] problem. Unfortunately, algorithms for the constrained problem are not yet commonly implemented in customary statistics packages. In fact, the same issue arises for the partial reaction orders of the reactants if the reaction network of interest contains different reaction pathways with drastically different kinetics. If, for example, the decomposition of A would be a second order process, e.g. due a self activation of the reactant being necessary, we would require two different transformation parameters for the concentration of A. In that case, as discussed before, the LASSO would need to be constrained to ensure disentangled results for the two reaction products. A workaround is to fit two independent models for both product species. We perform this individual analysis of the two target rates based on one data set. For this set we vary not only the feed concentrations but also the temperature to enable an estimation the activation barrier. Sampling these three factors in a 3-level FFD, we end up with 27 data points, cf. SI. Having one additional input feature, the rescaled temperatures, we increase the order of ansatz terms in this analysis to three. For both target variables, the rate towards P and the rate towards D, the algorithm produces a (now 2D) pareto front and a final solution can be selected as before. The resulting model parameters for these individual solutions are given in Table 3.3. The activation barrier for the product formation is identified to be 6.60 kcal/mol, corresponding to the barrier of the rate limiting decomposition of the activated catalyst complex. This is in nice agreement with the barrier for this step in the microkinetic model of 6.00 kcal/mol. Also for the side reaction, an activation barrier of 5.10 kcal/mol is obtained, in accordance with the reference value of 5.00 kcal/mol.

Effective Temperature Regimes for the CO Oxidation on Pt In industrial applications, the question frequently arises under which conditions the regime of the reaction system may change. This is important for the optimization of process and reactor parameters. To this end, our robust modeling approach can provide valuable insight. As previously shown for the HBr and model catalytic cycle examples, the sensitivity parameters (apparent reaction orders) change from one regime to another. Additionally, however, also the LASSO terms of the overall model can vary. This gives the algorithm enough flexibility to only select the limiting terms for the kinetic model of each respective regime. For CO oxidation on Pt, we know that there is a drastic change in surface coverage at around 500 K from a partially CO covered to a purely oxygen covered catalyst [106–108]. To identify potential temperature

regimes we need to determine effective kinetic models over a wide range of temperatures and analyze how the resulting models change qualitatively in various regimes. This comes with the advantage, that we do not need any prior knowledge about the microkinetic mechanism underlying the observed rates. All we need to specify is the variable input factors and the range of parameter values we are interested in.

It is especially important though to properly design the experiments for such a scan over a wider temperature range, as the goal is to identify relevant regimes with as few data points as possible. As an example, we generate kinetic data for the CO oxidation between 300 K and 700 K based on experimental designs [109] translated along the temperature axis. Details on the construction of these designs are given in the SI. Using this data set we can cut out smaller temperature ranges and fit local surrogate models. For reference, we create a large data set based on a FFD at every temperature with a total of 576 data points. Smaller partially overlapping subsets of 108 data points each, are cut out to determine local models for every temperature range. These local models are constructed using the proposed algorithm. An expansion of the polynomial terms up to 3rd order is used, as we need to take into account four input features, the concentrations of three species (CO, O_2 and CO_2), as well as temperature. Such a procedure results in a large number of pareto fronts, from which one solution needs to be selected respectively. This mandates an automated way of selecting the ideal solution to reduce on the one hand the required manual work, but on the other hand also to ensure more comparable and bias-free results. Here, we employ a rudimentary heuristic based on the same discontinuity arguments as before, which would guide a manual selection process. Details regarding this heuristic are explained in the SI. By observing the LASSO coefficients of these selected solutions we obtain a map, showing which model contributions are dominant at a given temperature. Figure 3.9 shows temperature maps for the CO oxidation on Pt. The resulting map based on the full factorial reference is given in Figure 3.9 c). It can be clearly seen, that the dominant LASSO terms undergo a drastic change over the range from 300 K to 700 K. Accordingly, we can roughly partition this map into four regimes and link these to characteristic changes in the underlying microkinetic model. To rationalize these findings, in Figure 3.9 a) we show the temperature dependence of the dominating rate constants from the microkinetic model. These rate constants correspond to elementary processes of the catalytic process, namely the adsorption of O and CO (ads_O, ads_CO), the desorption of CO (des_CO) as well as the surface reaction between the adsorbed species (reac). As further reference Figure 3.9 b) shows the concentration of adsorbed CO, resolved along both the temperature and the reactor axis.

Regime 1 (300 K to 420 K): At low temperatures up to about 420 K the interaction term $\{\hat{T}\}\{CO\}$ is dominating the models. This corresponds to the interaction of the temperature effect and the concentration of CO. Looking at the behavior of the rate constants of the elementary processes in Figure 3.9 a), we see that in this same low temperature range, the rate constant of the surface reaction between the adsorbed species (reac) changes dramatically, explaining the significant sensitivity to temperature. Furthermore, over the entire temperature range, the adsorption of oxygen is at least three times faster compared to the adsorption of CO, making CO the limiting species for the formation of carbon dioxide.

Regime 2 (420 K to 540 K): In the subsequent transition region from 420 K to 540 K the temperature influence is less pronounced. The dominating terms now contain mostly the concentrations of oxygen ($\{O_2\}$) and CO ($\{CO\}$). Looking at the rate constants (Figure 3.9 a)) there is no significant change to observe in this temperature range. The CO coverage (Figure 3.9 b)), on the other hand, undergoes a drastic change in this regime, from a partially CO covered surface to an approximately CO free catalyst above 530 K. For this reason, the relative contributions of oxygen and CO to the overall effective kinetic behavior do heavily depend on the temperature. This may explain, why the algorithm is not able to reduce the effective kinetics to a single dominating term for this range of temperatures.

Regime 3: (540 K to 600 K): The third regime again shows a stronger temperature dependence. The three body interaction ($\{\hat{T}\}\{O_2\}\{CO\}$) is the dominant term in this region. A similar reasoning as in regime 1 also applies here. The drastic change in the relative rate of CO adsorption and desorption leads to the observed temperature dependence.

Regime 4: (600 K to 700 K): The high temperature regime above 600 K is again dominated by effects of reactant concentrations ({O_2}{CO}). Temperature contributions are not significant and the rate is basically only limited by the small fraction of adsorbed CO on the mostly oxygen covered surface (cf. Figure 3.9 b)). This coverage of CO highly depends on both the gas phase concentrations of CO and oxygen, as in the underlying microkinetic model both species compete for the same adsorption sites.

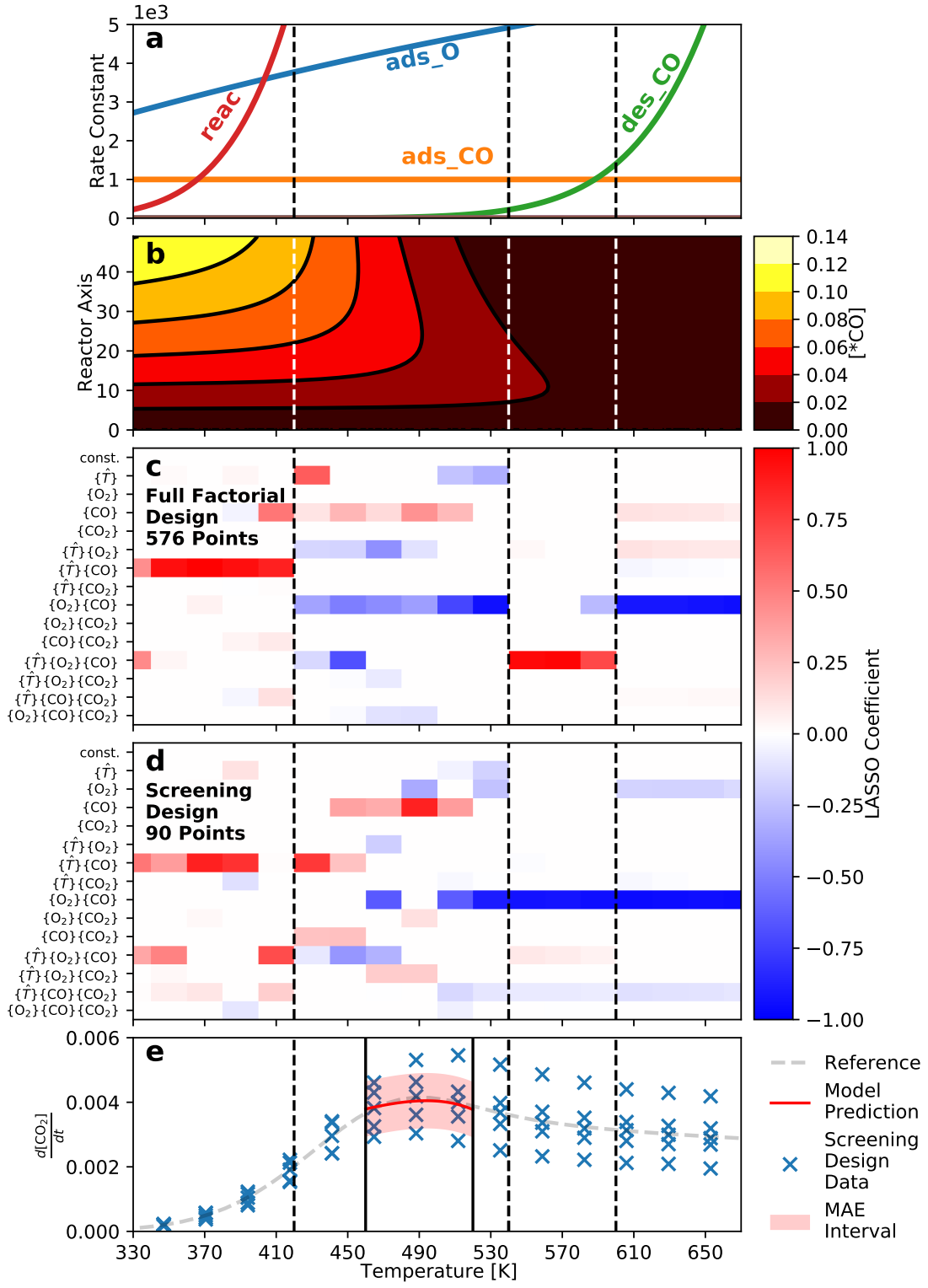


Figure 3.9: (Caption next page.)

Figure 3.9: (Previous page.) Temperature scans for the CO oxidation model. a) Temperature dependence of the dominant rate constants from the underlying kinetic model (Details provided in the SI). b) Concentration of adsorbed CO species on the catalyst sites versus temperature from the microkinetic model. While there is some CO coverage present at lower temperatures, above 570 K there is basically no adsorbed CO left. c) Reference Temperature map based on a full factorial experimental design with a total of 576 points. d) Temperature map based on a set of screening designs with a total of 90 data points. e) Rate data versus temperature for the entire range. The blue markers refer to all simulated rates at different conditions within the 90 point screening design. The grey dashed line gives the ideal rate at constant feed composition ([CO] = [O_2] = [CO_2] = 1 mol/l) and acts a guide to the eye. The red line and area shows the fitted model prediction in the regime between 460K and 520K with the mean absolute error (MAE) interval.

In reality it is, of course, not possible to experimentally measure almost 600 concentration profiles for such an analysis. Therefore, we require a means of more efficiently sampling the relevant data. In Figure 3.9 d) we show the temperature map obtained based on a combination of low-resolution screening designs with a total of only 90 profiles. The corresponding subsets consist of only 15 data points at three temperatures each. This extremely small amount of data challenges the estimation of the PPCC values. Still, as we are mostly interested in a rough estimate of the change in LASSO coefficients, such screening designs can be applied. In order to fit an accurate surrogate model, applicable for the prediction of reaction rates, a larger data set will be required. Looking at Figure 3.9 d), we can see that the qualitative features of the map are very similar to the full factorial reference in Figure 3.9 c). Regimes 1 (300 K - 420 K) and 2 (420 K - 540 K) are well captured by the screening designs, and also the high-temperature behavior is reproduced. The main difference between the two designs occurs in regime 3 (540 K - 600 K). Here, the temperature contribution is missing and the dominant term is {O_2}{CO} instead of {$\hat{T}$}{O_2}{CO}. Still, there is a small contribution by {$\hat{T}$}{O_2}{CO} also present for the screening design. This shows, that an initial assessment of relevant temperature regimes can already be obtained with a much smaller amount of data, while a detailed modeling of the most interesting regimes could require additional sampling in a second step.

Assuming that we do not have access to the extensive full factorial data, we would identify three temperature regimes based on the temperature screening map. In addition to that, we obtain a rough estimate on how the reaction rate behaves with changing temperature (blue crosses in Figure 3.9 e)). It turns out, that the reaction rate reaches its maximum in the range around 500 K. Looking at the temperature map, this maximum region is located within regime 2. As mentioned before, regime 2 does not show a clear dominant model term, which can be attributed to the constant change in coverage in this temperature range. The maximum in reaction rate, however, can further explain the appearance of multiple model terms in the temperature map. To form such an extreme point in terms of the rate, at least two counteracting processes need to take place in parallel. These processes are represented by different LASSO terms with their relative importance determined by the coverages of the reactants. In order to more accurately model this region of interest, additional data points are acquired in the range between 460K and 520K based on a 3-level full factorial design yielding 81 additional points. From the screening design data we know, that the maximum rate should be located in this region. Using this more extensive data set we can fit a surrogate model for this entire temperature range. In this model, also quadratic terms are included to better describe the extremum. We have discussed before, that purely quadratic and higher power terms are usually neglected in our model ansatz. However, if we know, that there is an extremum located within our region of interest, including quadratic terms can be beneficial. The red line in Figure 3.9 e) indicates the model prediction for the reaction rate at constant feed composition ([CO] = [O_2] = [CO_2] = 1 mol/l) with an area showing the mean absolute error (MAE) of the model based on the training data. Using the additional data points, our algorithm is able to accurately reproduce the curvature of the reaction rate as a function of temperature in this region. The reference data (grey dashed line) for the rate at the same constant feed composition falls well within the error interval. The model predictions here are only given for the region of interest between 460 K and 520 K. An extrapolation to both higher and lower temperatures is possible up to a certain degree. However, while

polynomial models excel at locally approximating complex functions, one has to be careful with extrapolations beyond the training region. Extrapolation based on the presented model is presented in the SI. In Figure 3.10 additional details on the fitted kinetic model for the maximum rate region are given.

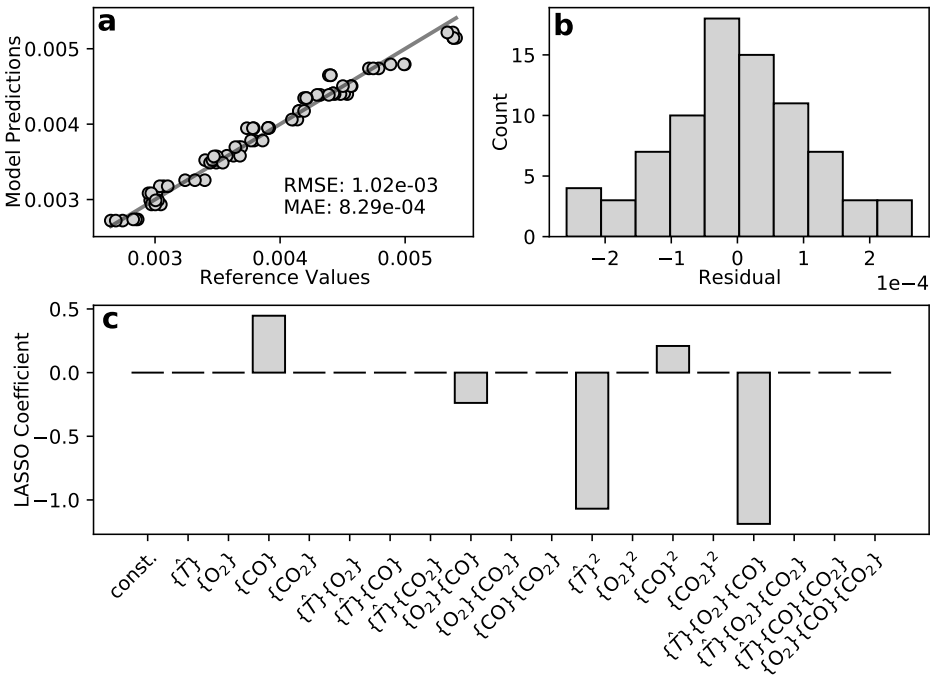


Figure 3.10: Details for the maximum region (460 K to 520 K) effective kinetic model identified by the screening approach of Figure 3.9 b): a) Parity plot of the model prediction versus the reference from the microkinetic model. b) Residual distribution. c) LASSO coefficients for the maximum region model.

Inspecting the LASSO coefficients (e.g. the selected model terms) in Figure 3.10 c), we see that in nice agreement with the temperature map, multiple terms contribute to the final model. The two largest contributions come from the 3-body interaction ($\{\hat{T}\}\{O_2\}\{CO\}$) and the quadratic term ($\{\hat{T}\}^2$). The former indicates a strong interplay between CO and O_2 with a significant temperature dependence. Coming back to the strong changes in surface coverages in this temperature region, this is to be expected. The large negative coefficient of the quadratic temperature term on the other hand is in nice accordance with the maximum in rate observed in Figure 3.9 e). The remaining model terms all depend on the concentration of CO ($\{CO\}$), with two of them ($\{CO\}$ and $\{CO\}^2$) showing a positive coefficient. This points at the growing dependence of the overall rate on the concentration of CO at higher temperatures as also observed for the high temperature regimes in the temperature map. Finally, we need to analyze the sensitivity parameters (apparent reaction orders and barrier) determined by the algorithm (Table 3.4). The apparent activation barrier of around 7.600 kcal/mol is in reasonable agreement with the barrier for the surface reaction in the microkinetic model of 10.000 kcal/mol. A negative reaction order for O_2 indicates a detrimental effect of an increase in oxygen concentration on the reaction rate. We can understand this, as the underlying microkinetic model does only consider one type of adsorption site. Therefore, an increase of oxygen concentration may lead to a reduced adsorption rate of CO. At the same time, an increase in CO concentration will lead to an increase in reaction rate causing the positive reaction order in CO. The concentration of CO_2 does not influence the reaction rate, as no readsorption of the product species is considered. This leads to a reaction order of effectively zero in the concentration of CO_2. This detailed analysis of the maximum rate region (460 K to 520 K) shows, that our algorithm can not only be used to identify interesting kinetic regimes of an unknown system with very limited data. It can also reproduce the effective kinetics of a local regime in a subsequent step. Further, we can interpret the resulting

Table 3.4: Model parameters and objective values for the selected solution for the maximum region of the CO oxidation example. The corresponding pareto front is given in the SI.

$\alpha_{\hat{T}}$	$\alpha_{[O_2]}$	$\alpha_{[CO]}$	$\alpha_{[CO_2]}$	Residual	PPCC
0.760*	-0.610	0.306	0.004	1.346	0.997

* in 10 kcal/mol

model function in a physically meaningful way. Of course, within this study, the physical meaning of the obtained models is limited by the assumptions of the underlying simulated reaction models.

Going beyond the pure interpretation of the model parameters and terms, we can also use the obtained models in an optimization process to identify reaction conditions which maximize the rate. For more complex reaction networks such an optimization could even target other properties like selectivities towards certain products. A closer look at the predictions along cuts through all dimensions of the rate model (the three feed concentrations and temperature) shows that our model does indeed not only describe the temperature trend but also the effect of the feed composition and could therefore be used in such an optimization.

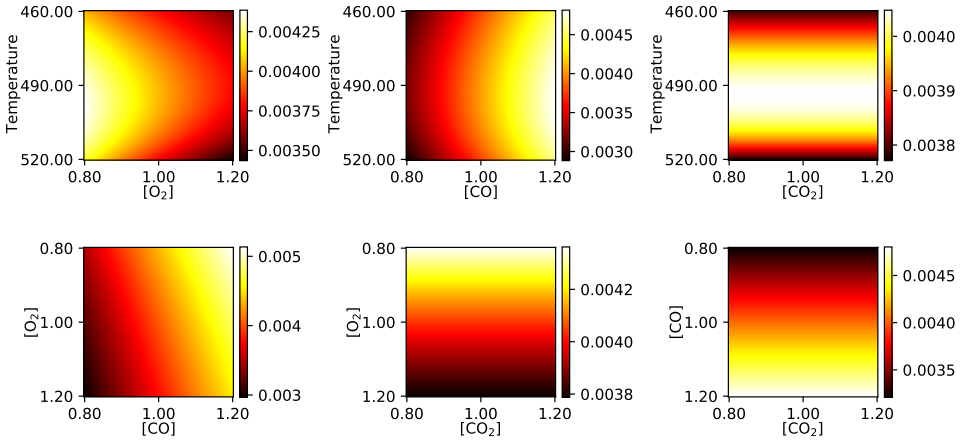


Figure 3.11: Maximum region model predictions for 2D cuts through the 4D model. The remaining two factors are kept constant at their respective central values (1.00 for [X] and 490 K for T).

The 2D cuts shown in Figure 3.11, reproduce the behavior shown for the temperature scan designs with a clear maximum around 490 K to 500 K. The most interesting cut is the one through the [CO] [O_2] plane in the lower left of Figure 3.11. This contour plot shows, that the model also captures the influence of the CO and oxygen feed concentrations and would predict a higher reaction rate, if the feed contains more CO and a lower amount of oxygen. These trends are perfectly in line with the obtained effective reaction orders.

3.1.5 Outlook: Application to Experimental Data

We presented a novel approach to the analysis of reaction kinetic data with the goal of not finding the most accurate model, but the most robust one. By analyzing both the residual norm and the distribution of errors for potential candidate solutions, we propose an algorithm to identify the optimal trade-off. With the main goal of reducing the amount of systematic error through model assumptions, we let the algorithm choose polynomial terms making use of the sparsifying LASSO regression method. Additionally, the algorithm determines sensitivity parameters which can be interpreted as physical quantities like activation barriers or apparent reaction orders. Besides yielding rate law like models, the algorithm

can further provide valuable insight into the underlying processes. This way, a meaningful analysis of reaction kinetics can be performed without knowledge of the exact reaction mechanism. Emphasizing the aspect of robustness in such an analysis may become especially relevant with respect to transferability of the obtained models.

We showed that such a procedure reproduces well understood reaction systems like the hydrogen bromine reaction or the catalytic model cycle without prior knowledge. In the HBr case, the empirically observed fractional reaction order of 3/2 was automatically detected by our method. At the same time we obtained an interpretable rate law capturing the effect of not only the reactant concentration but also the influence of temperature. The homogeneous catalytic cycle was equally well described and we were able to capture the change in effective kinetics with an increasing excess of one reactant by monitoring the sensitivity parameters resulting from the automatic analysis. Going one step further towards industrially relevant processes, an analysis of the CO oxidation kinetics on Pt over a wide temperature range provided both, information about process regimes, as well as an accurate surrogate model for the region of interest. While these surrogate models can not provide detailed knowledge about the reaction mechanism, valuable indications can be gathered by a thorough analysis of both the transformation parameters (α) and the polynomial terms selected by LASSO. This way we can learn how reaction orders and limiting species may change with reaction conditions like the feed composition.

It is important to mention, that this entire procedure relies on rationally designed experimental data sets, something not frequently encountered even in recent studies. A factorial sampling of experimental data, varying multiple factors at the same time is vital for a meaningful inference on interaction terms. As known from chemical reaction kinetics and also shown by our results, these interactions terms are most often the key contributions to the rate laws of various reactions. The artificial data sets used in this work rely on classical experimental designs. For the CO oxidation example, making use of a simple construction based on established screening designs we were able to achieve accurate results in a data efficient way. In the future, the data requirements for such an analysis could even further decrease by employing especially designed optimal designs [110, 111] for such specific use cases. Applying this algorithm to more complex reaction systems and especially working on real experimental data will be necessary to benchmark against established methodologies. Experimentally obtaining suitable training data for such an analysis will be a challenge in its own right. The algorithm requires kinetic data reflecting the chemical reaction kinetics. Therefore, such experiments need to be performed in specialized reaction sensitive reactor setups [112–114]which are not dominated by transport phenomena. While a polynomial approximation of the data in a transport limited reactor could in principle yield accurate results, any kind of interpretability in terms of kinetic parameters is lost in such cases. Provided suitable training data, however, our approach promises to provide efficient kinetic models which, coupled to different transport models, should even be transferable between different reactor geometries. This would allow for a straight forward upscaling going from idealized lab scale experiments to real industrial reactors. Obviously, our method is not designed to describe the global kinetics of a system over a wide range of conditions. The resulting models are also not intended to describe transport phenomena in chemical reactors. This approach rather resembles a semi local sensitivity analysis, which provides simple and well behaved kinetic models.

The formulation of this method allows for a straightforward extension in terms of both input and response variables to deal with more extensive reaction networks. In addition to that, also the optimization cost function may easily be modified to employ other error models for specific use cases. In this context, it may also be useful to apply our algorithm to data from complex microkinetic models. Comparing the result of such an analysis on experimental data versus model predictions, both in terms of sensitivity parameters and dominant LASSO terms, could be an extremely sensitive measure of the quality of a microkinetic model. Further, the capability of our approach to obtain robust effective kinetic models without any prior knowledge of the system opens up a wide field of possibilities towards computer guided experimentation. By optimizing desired properties like turn over frequencies or selectivities on the level of the cheap effective kinetic models, an algorithm can propose potentially interesting parameter ranges to sample next in an iterative procedure.

This way, a fully automated exploration of relevant regimes as well as optimization of the process conditions could be realized.

3.2 Surrogate Models for Multi-Regime Kinetics

This chapter builds on the capability of the previously established sparse polynomial approximation models to differentiate between different kinetic regimes based on the effective behavior of local models. A general, largely automatized regime identification algorithm is presented, which utilizes the local model parameters as a fingerprint to localize kinetic phase transitions. The first section of the chapter outlines the main steps of the regime identification algorithm. Following this exposition, the CO oxidation on Pt example from the previous chapter is continued to illustrate how to couple local effective models into a global representation over multiple regimes. Investigating a MKM for the CO oxidation on RuO_2, the last two sections describe two and three dimensional problems.

3.2.1 Introduction

Reaction kinetic systems often show dramatically different behavior depending on the conditions at which the reaction is performed. In general, such a change in effective behavior is not continuous over a wide range of reaction conditions but rather happens at a specific critical point, thus breaking the underlying smoothness assumptions of many machine learning approaches. These so called kinetic phase transitions can for example be observed for catalytic surface reactions in both theory and experiment [34, 115–117]. For such systems, we can segment the reaction condition space into different regimes, where the effective behavior of the reaction does not change qualitatively and the, often sharp, boundaries between those regimes. In many examples, such phase transitions are correlated with the surface states of the catalyst and for example poisoning of the catalytic surface can cause a dramatic drop in the reaction rate [117]. An extensively studied catalytic system of this type is the CO oxidation reaction over RuO_2. In this system, several sharp phase transitions can be observed as a function of the partial pressures of CO and O_2 [3]. We can set up microkinetic models, which reproduce these kinetic phase diagrams over a wide range of reaction conditions. For this specific system, it was shown that even a comparably simple mean field MKM is capable of capturing most phase transitions [12]. Unfortunately, knowledge of the microscopic reaction mechanism is a strict requirement for setting up such a MKM. In the previous chapter we could show that the sparse approximation approach we propose is capable of fitting effective models similar to the rate laws of classical kinetics. These rate laws correspond to an approximation to the dominant effective kinetics in a given reaction condition range and can be obtained without a detailed understanding of the reaction system. Therefore, the range of validity of such a rate law should provide a good estimate at which critical conditions the behavior of the system changes. The regime identification algorithm we propose in this work locally probes the kinetics of the system to determine the position of kinetic phase transitions. Local rate law kinetic models provide information on which experimental conditions contribute to the kinetic model as well as on the corresponding nonlinear sensitivity parameters. Using this information as local features we can sample the reaction condition space and differentiate between regions of distinct kinetic behavior. As soon as the kinetic phase transition is localized, we can couple the locally valid effective rate laws via a transition function across the phase boundary. Making use of clustering analysis and classification models, we implement this procedure to automatically identify kinetic regimes in order to create a global surrogate model by coupling multiple local rate law like kinetic models.

3.2.2 Regime Decomposition Algorithm

Here, the main steps of the regime identification procedure are outlined:

Sampling the Reaction Parameter Space First of all, we need to define the global bounds of the reaction conditions we want to describe. Having settled on some finite space, we generate a sequence of Quasi Monte Carlo (QMC) [68] points to sample this space as evenly

as possible. Every one of these QMC points is then the center of a classical experimental design, which gives a more detailed factorial description of the local environment of these centers. These local designs (LD) are used as the training data for the local analysis in the subsequent step.

Local Kinetic Models For the local approximation of the reaction kinetics we use the data driven modeling methodology described in the previous section. This method yields sparse polynomial models in transformed variables (SPTVs). Such an SPTV resembles an approximation to the rate laws known from classical kinetics and provides us with interpretable kinetic parameters like reaction orders or apparent barriers. Further, these polynomial models are extremely cheap to evaluate, smooth and continuously differentiable. Another valuable feature of the SPTV is its capability to identify dominant terms in different kinetic regimes making use of the LASSO [1, 6] regression to select only relevant model terms for the final local model. We have previously shown, that such an analysis can differentiate between multiple effective temperature regimes in the catalytic oxidation of CO over Pt(111). Performing this local analysis, we end up with one SPTV model for every QMC center point. An example for the functional form of such an SPTV is given in equation 3.14 for a system with two relevant process parameters x_1 and x_2:

$$\eta(\boldsymbol{x}, \theta, \boldsymbol{\alpha}, \beta) = \left(\theta_0 + \theta_1 \cdot x_1^{\alpha_1} + \theta_2 \cdot x_2^{\alpha_2} + \theta_{12} \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2} + \theta_{11} \cdot x_1^{2\alpha_1} + \theta_{22} \cdot x_2^{2\alpha_2} + ...\right)^{\beta}, \quad (3.14)$$

where the reaction rate model η is a polynomial function of the reaction parameters $\boldsymbol{x}_i$ like e.g. feed concentrations and the α_i are the corresponding sensitivity parameters. The θ_i and θ_{ij} are the regression coefficients of the various polynomial terms. Due to the sparsifying LASSO regularization, in practice, only few of these coefficients are nonzero, resulting in compact and interpretable models.

Regime Identification In order to automatize the decomposition of the global domain into multiple regimes, we employ clustering analysis (c.f. section 2.9). Clustering is considered an unsupervised learning method to identify structure in a data set. Data points are sorted into different domains based on a cost function like for example a proximity measure. The goal here is, to decompose the global range of reaction conditions into regimes of different effective kinetic behavior. Therefore, we want to identify which local models have a similar form, i.e. describe a similar dominant kinetic process. The data set, we want to perform the cluster analysis on, can be constructed from different model parameters like on the one hand, the LASSO coefficients of the individual local SPTV models, and on the other hand, their respective transformation (or sensitivity) parameters. In some cases, in addition to that, we can also consider the input reaction parameters in the clustering to ensure compact domains with low overlap. The specific clustering algorithm we use is the *Kmeans* algorithm (c.f. section 2.9.1). The optimal number of domains is evaluated based on a scree plot of the clustering inertia versus k. The final value for k is then either selected manually or based on the numerically identified point of highest curvature of the scree plot. Other clustering methods like the density based method "Ordering Points To Identify the Clustering Structure" (OPTICS) [118] have the advantage that the number of clusters does not need to be predefined. For small data applications, in general hierarchical clustering methods are recommended, however, in this study we applied the *Kmeans* algorithm due to its conceptual simplicity and interpretability [119]. Having performed this step, we can assign a clustering label to every QMC point, which allows us to decompose the reaction parameter space into local domains. As the parameter space in which we perform the clustering is generally more than three dimensional, we use principal component analysis (PCA) [41] as a tool to visualize the resulting clustering. PCA is a tool to decompose multidimensional data into orthogonal linear combinations which describe the maximum amount of variance within the data. This way, assuming linearity, we can assure a minimal loss of information projecting onto a lower dimensional representation.

Coupling Local Effective Models In order to describe the entire global domain of reaction conditions, we want to have a single model capturing the reaction kinetics over multiple

regimes. Having identified multiple reaction regimes within this global domain, we can, in a first step, reduce the number of local models to the necessary minimum. For every local regime, we collect all QMC centers, which have been assigned to this regime in the clustering analysis. Based on these regime wide data sets, we, again, fit a SPTV model which should now be valid for the entire regime. The only question left, is now, how to couple these regime models. This is done using a support vector classification model (SVC) [9]. The SVC model is trained on the reaction conditions as an input to reproduce the clustering labels obtained by *Kmeans*. This way, the SVC can predict in which regime a certain point is located, provided a set of reaction parameters. The most interesting regions of the global domains, are, however, the boundaries between different local regimes. In order to describe these boundaries in a somewhat smooth manner, we use the SVC to estimate the membership probabilities of a point (set of parameters) for every identified regime using a cross validation technique [76, 77]. With these membership probabilities as weights, we can formulate our global surrogate model as a linear combination of SPTV models. The global surrogate model can be written as

$$\hat{\eta}(\boldsymbol{x}, \hat{\boldsymbol{\theta}}) = \sum_k p_k(\boldsymbol{x}) \eta_k(\boldsymbol{x}, \theta_k, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) \tag{3.15}$$

where the estimated rates $\hat{\eta}$ at a point $\mathbf{x}$ are given by the sum over the membership probabilities p_k multiplied by the predicted contributions of the regime specific SPTV models η_k. A schematic of the entire regime identification procedure is shown in Figure 3.12.

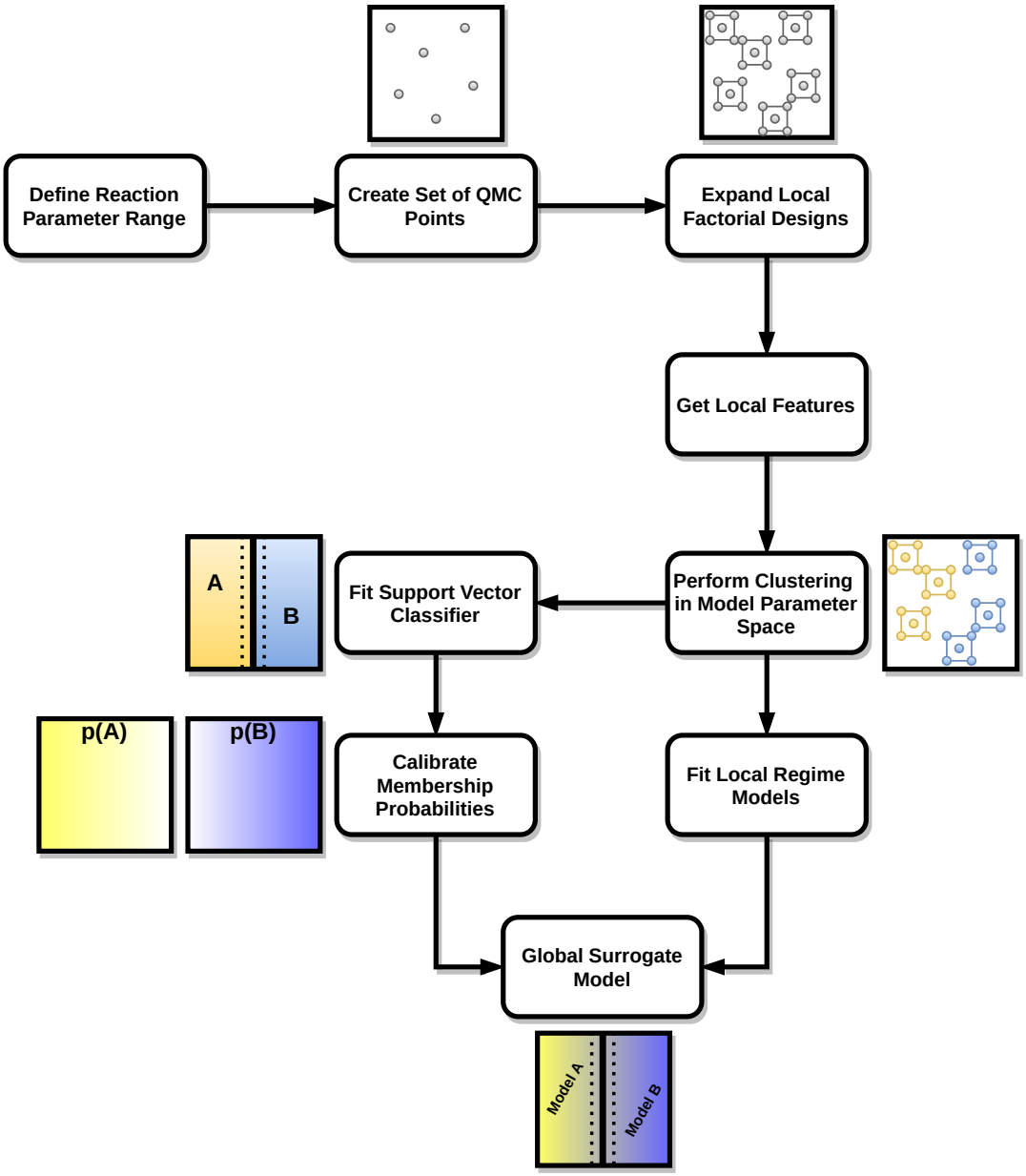


Figure 3.12: Schematic of the regime identification procedure. Starting with an initial sampling of the global domain, local regimes are identified based on the similarities of the local models identified for the respective QMC points. The predicted membership probabilities from the calibrated classification model are then used as weights for coupling the local regime models.

Implementation Details Building on the Python implementation of the SPTV modeling algorithm, the regime identification algorithm is also written in Python. Further, the scikit-learn [99] implementations of the *Kmeans* algorithm and the SVM (which also provide the probability calibration methods) are used. The QMC samples are constructed using the scipy statistics package [100].

3.2.3 Proof of Concept: CO Oxidation on Pt

Before we apply the regime decomposition procedure to a more complex microkinetic model, we want to briefly revisit the CO oxidation on Pt example from the previous section 3.1.4. By analyzing the change in model parameters over a wide temperature range, we were able to identify temperature regimes with distinct apparent kinetics. For each of these regimes, an individual kinetic model could be identified. Using only a single SPTV model, we would

not be able to capture the entire range of reaction conditions. However, as the LASSO coefficients within the four major temperature regimes did not show too drastic variations, a set of coupled local models should be able to achieve this. Using the same screening design data set as discussed in section 3.1.4, we apply a clustering analysis to the set of local sensitivity parameters (α) of the SPTV models for every temperature interval. In the same way as the overall temperature range was segmented into effective regimes by visual analysis in the previous chapter, temperature intervals with similar effective kinetics should now also be assigned to one cluster. The LASSO coefficients are not included in the model parameter data, as due to the almost smooth transition between the low and high temperature regimes (c.f. Figure 3.9 d) these parameters do not allow for a clear assignment of the data points to the transition region. The results of the *Kmeans* analysis are presented in Figure 3.13.

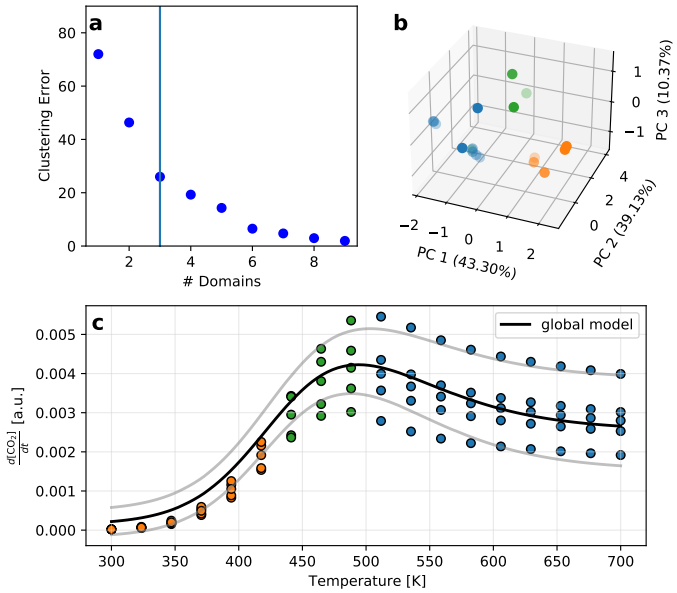


Figure 3.13: a) Scree plot for the clustering of the local sensitivity parameters α for the CO oxidation on Pt. The vertical line indicates the chosen number of domains $k = 3$. b) PCA representation of the sensitivity parameter space. The coloring of the points resemble their respective assignment to a cluster and the percentage numbers in the axis labels refer to the amount of explained variance by each principal component (PC). c) Screening design data colored to match the identified clustering of the local models. The black line corresponds to the reaction rate predictions of the SVC coupled global model at constant feed composition ([CO] = [O_2] = [CO_2] = 1 mol/l). The spread of the underlying data points comes from the influence of the feed composition. To illustrate that this behavior is also captured by our global model, the gray lines indicate the model predictions for two different feed compositions, a CO rich ([CO] = 1.2 mol/l, [O_2] = [CO_2] = 0.8 mol/l) and a oxygen rich ([CO] = [CO_2] = 0.8 mol/l, [O_2] = 1.2 mol/l) composition.

The scree plot in Figure 3.13 a) indicates that going beyond three domains, only small improvement in terms of the clustering error can be gained. Choosing three clusters for the *Kmeans* analysis leads to a reasonable segmentation of the samples in terms of their respective sensitivity parameters as can be seen in the PCA representation 3.13 b). Thus, the sensitivity parameters yield a clear segmentation into three regimes (clusters). Again, for each individual regime, a SPTV model can be set up to represent the local kinetic behavior. In order to couple these regime wide models into a globally valid representation, we train a SVC model, which classifies points into individual regimes with respect to their coordinates in factor space (feed composition and temperature). In Figure 3.13 c) the screening design data points have been color coded according to their cluster label predicted by the SVC. We can see that this labeling results in three compact regimes along the temperature axis. Using the predicted membership probabilities of this classification model, we can couple the local kinetic models for the three identified regimes in a linear combination. The resulting global model is shown in Figure 3.13 c) (black line). The combination of three local SPTV

models yields an accurate description of the reaction rate over the entire temperature range, including the transition region with the extremum. Following the same ideas, in the next section, we will apply this procedure to a more complex two dimensional example.

3.2.4 Test Case: Isothermal CO Oxidation on RuO_2

For a second test case, we analyzed a microkinetic model for the CO oxidation on a RuO_2 catalyst (c.f. 2.11.4) using the previously explained regime identification methodology. In this example, the complex reference model is a mean field MKM as presented in section 2.11. We define the global reaction condition domain as a partial pressure range of $10 \cdot 10^{-2}$ to $10 \cdot 10^{2}$ bar for both gas phase reactants CO and O_2. The sampling of this range is performed using a set of 64 QMC points drawn from a Halton series, scaled to the respective bounds. Around each of these QMC points for the local design (LD) a four level full factorial is expanded. The QMC point is defined as the central factor configuration (0,0) of the LD to which all displacements are referenced. The magnitude of the maximum displacements are determined as one third of the averaged nearest neighbor distance of the QMC points, resulting in a range of about 0.125 in terms of the logarithm of both partial pressures. This way, we ensure an approximately even sampling with minimal overlap. Each of these LDs consists of 16 point respectively, hence the training data for each local SPTV model consists of 17 points, including the QMC center. This results in a total of 1088 evaluations of the MKM. The distribution of the QMC central points and the local designs over the reaction condition range are illustrated in Figure 3.14.

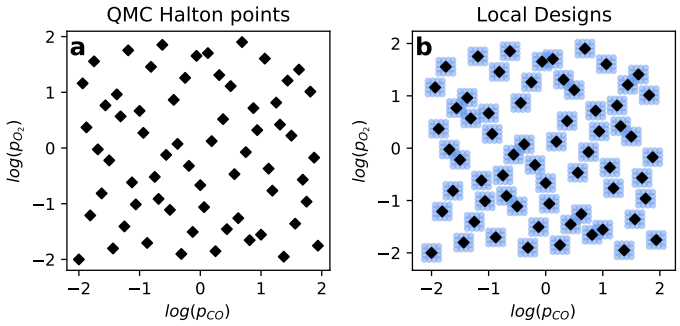


Figure 3.14: a) Distribution of QMC points over the reaction conditions space. b) Local factorial designs constructed around every QMC center. The maximum displacements of the local designs are one third of the averaged QMC point nearest neighbor distance: $d(\log(p_{CO})) = d(\log(p_{O_2})) = 0.125$.

For the local models, the polynomial ansatz is restricted to only linear terms, as these models purely serve as a qualitative indication of the locally dominant factors. It is important to mention that, in order to ensure numerical stability, these SPTV models map between the logarithmized quantities $\log(p_{CO})$, $\log(p_{O_2})$ and $\log(\text{TOF})$. Therefore, the resulting sensitivity parameters may not directly be interpreted in terms of partial reaction orders. As explained in the previous section, the model selection from the pareto front is performed automatically.

Surface Phase Diagram In order to rationalize the results of our regime identification formalism, we inspect the system at hand in more detail: The dependence of the MKM predicted TOF for CO oxidation is shown in Figure 3.15 a). We can see that there are dramatic differences in TOF depending on the partial pressures of both reactants. The region of maximum activity is visible as a diagonal stripe with log(TOF) values up to 6. This region is characterized by a combination of moderate partial pressures of both CO and O_2. The opposite would be the regions in the upper left and lower right corners of Figure 3.15 a), in which one of the two partial pressures is orders of magnitude higher than the other.

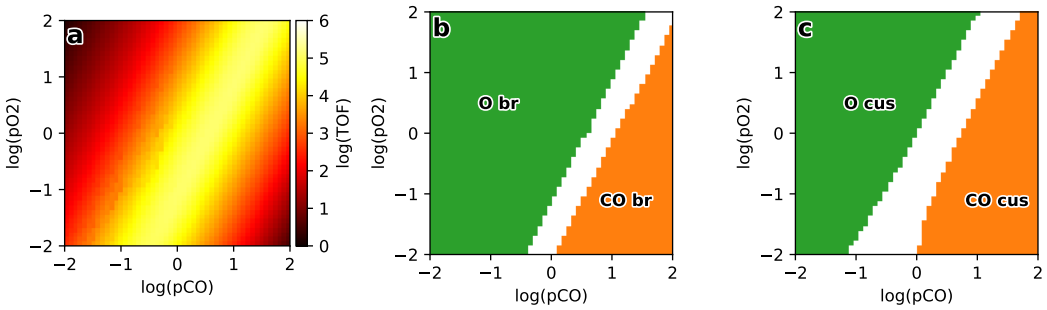


Figure 3.15: a) TOF predicted by the mean field MKM. b) Surface phase diagram for the bridge type adsorption sites (br). c) Surface phase diagram for the cus type sites. The colored areas indicate that one surface species is dominating with a coverage $\geq$ 90 %. White areas correspond to conditions for which no single species shows such a high coverage.

Comparing the TOF to the surface phase diagrams in Figure 3.15 b) and c), the correlation between surface coverage and activity directly becomes obvious. For both, bridge and cus adsorption sites, the surface coverage is dominated by a single species in most of the pressure range. At low p_{CO} and high p_{O_2}, the catalyst surface is almost entirely covered by atomic oxygen, while in the other extreme (low p_{O_2} and high p_{CO}) adsorbed CO occupies most adsorption sites. The intermediate region, where neither of the two adsorbed species is poisoning the catalyst, is the therefore of interest. This range also coincides with the maximum TOF region in Figure 3.15 a). This behavior can intuitively be understood, as the MKM mechanism dictates that both reactants need to be in an adsorbed state at the same time in order to react towards CO_2. Therefore, the goal of the regime identification would be for the algorithm to realize that the effective kinetics of the reaction are dominated by different terms in the two extrema of the surface phase diagrams. This separation of the overall reaction condition space directly allows for locating the intermediate pressure regime as the main region of interest.

Identified Kinetic Regimes With the parameter sets of all local SPTV models, we perform a clustering analysis in order to classify the QMC centers to regions of the same effective kinetics. The model parameter space is spanned by the two model sensitivity parameters ($\boldsymbol{\alpha}$) as well as the LASSO coefficients for the individual polynomial terms $\boldsymbol{\theta}$. For a linear model in two variables $\log(p_{CO})$ and $\log(p_{O_2})$ this results in a 5 dimensional space. A reduced 3 dimensional PCA representation of this parameter space is shown in Figure 3.16 b). Each point in this space represents one local SPTV model. The three principal components account for almost 90 % of the explained variance in this data set, thus only minor information is lost by displaying the parameter space in 3D. As we can see, the data points form three distinct clusters and the scree plot in Figure 3.16 a) further confirms $k = 3$ as a reasonable choice.

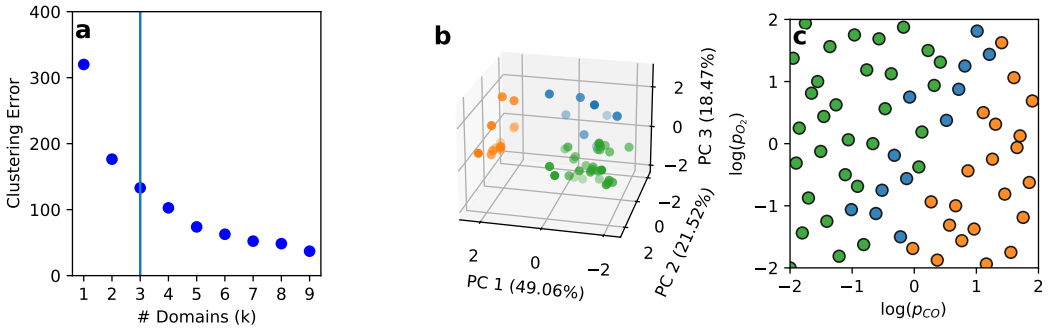


Figure 3.16: a) Scree-Plot for the *Kmeans* clustering of the model parameter space. The vertical line indicates the chosen number of clusters ($k = 3$). b) PCA representation of the model parameter space. The colors of the individual points indicate their clustering labels. The percentage of explained variance of the individual principal components is given in the axis labels. c) The center points represented in reaction condition space. The coloring, again, indicates the cluster assignment.

Having identified a clustering of the local models in parameter space, we need to verify that this classification is also meaningful in terms of the reaction conditions (i.e. the partial pressure ranges). Figure 3.16 c) shows the distribution of the QMC points in reaction condition space and the coloring of the points indicate their cluster assignment according to the *Kmeans* analysis. It immediately becomes obvious that the spatial compactness of the identified clusters is not only given in parameter space, but the clusters also form contiguous regimes in the reaction condition space. Further, the three identified clusters show a very similar structure to the catalyst surface phase diagrams. While there is some overlap between the point distributions of the green and blue clusters in Figure 3.16 c), this separation suggests the blue cluster to resemble the maximum TOF transition region, while the green and orange clusters correspond to the O and CO covered regimes, respectively. This becomes even more pronounced looking at the results of the support vector classification trained on the clustering labels. Figure 3.17 shows a direct comparison of the SVC domains with the surface phase diagrams from the MKM.

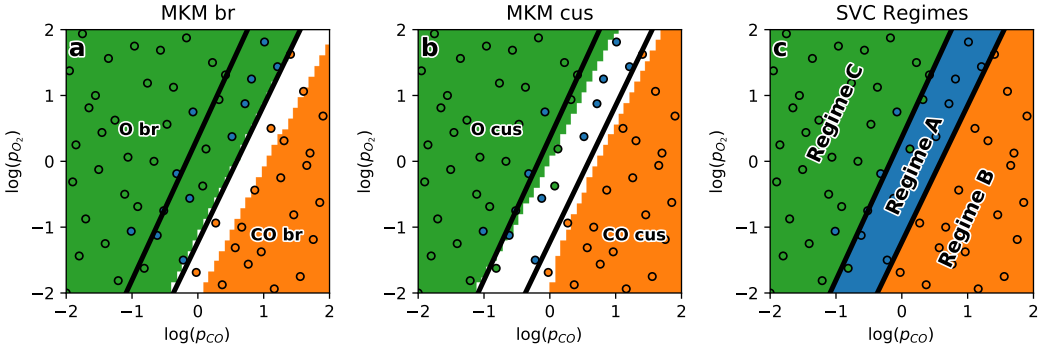


Figure 3.17: SVC predicted domains: a) Surface phase diagram of the bridge (br) sites. b) Surface phase diagram of the cus sites. c) Classification of the reaction condition space according to the SVC. The black lines indicate the SVC decision boundaries. The coloring of the points corresponds to their assignment to different clusters during the *Kmeans* analysis.

Overlaying the two decision boundaries of the SVC with the surface phase diagram for the cus adsorption sites like shown in Figure 3.17 b), we see a very nice agreement. Most points assigned to the blue cluster lie within the surface phase, which is not dominated by a single adsorbed species. The agreement between the SVC and the surface phase diagram of the bridge sites is lower compared to the cus sites. The SVC predicted transition region is slightly shifted towards lower CO partial pressures compared to the phase transition region. The qualitative dependence on the O_2 partial pressure, however, is reproduced. This could be an indication that within the here studied pressure ranges, the effective kinetics of the

reaction are dictated by the reactivity of the cus sites. Indeed, the MKM rate constant for CO_2 formation from reactants both adsorbed on bridge sites is significantly lower compared to reaction involving at least one cus site (cf. section 2.11). This sensitivity of the apparent kinetics with respect to the reaction on the cus sites has also been reported in literature [12, 120]. Using the Platt scaling formalism, we can use the SVC to predict calibrated membership probabilities for the three identified regimes. The resulting probability functions are shown in Figure 3.18.

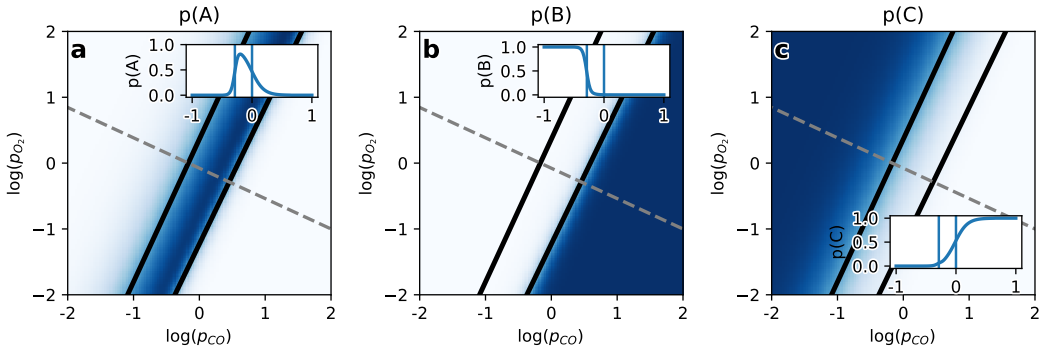


Figure 3.18: Membership probabilities calibrated using the extended Platt scaling formalism (cf. section 2.10). a) Membership probability for domain A. b) Membership probability for domain B. c) Membership probability for domain C. The insets show a line profile of the probability function orthogonal to the separating hyperplanes.

In general, these probabilities again show a clear distinction between the different regimes. A closer look at the insets in Figure 3.18, however, reveals that the two decision boundaries are not equally well defined. As was already visible in Figure 3.16 c), there is a overlap between the blue and green clusters. This is also resembled in broader probability transition between these two regimes as shown by the difference in width of the two sigmoid functions in the insets of Figure 3.16 b) and c), as well as the asymmetric tails of the probability function in Figure 3.16 a). We can see that these calibrated membership probabilities accurately reproduce the previously established separation of the reaction condition space and can thus be used as weights for the global surrogate model in the subsequent step.

Local Regime Models In order to set up the global surrogate model for the entire reaction condition space, we first need to fit the three individual regime models. To this end, for each regime, we combine the data from every QMC center assigned to it and fit another SPTV model for the joint data set. To more accurately describe the kinetics of these regimes and make efficient use of the large amount of data per regime, for these models we expand the polynomial terms up to second order. The model parameters for the three automatically selected pareto optimal models are given in Table 3.5. For regime A, the transition region, the absolute local sensitivity towards the partial pressure of O_2 is smaller than for CO, with CO showing a negative α value. Only two LASSO coefficients have been selected for this regime model, the interaction between the two reactant pressures and the quadratic contribution by O_2. The negative coefficient $\theta^2_{O_2}$ nicely shows, that we have an extremum (a maximum) in TOF within this region, while the large coefficient for the interaction term reproduces the almost diagonal structure of the regime in the reaction condition space. The insets in Figure 3.18 show the predicted membership probabilities across this phase transition along the line $\log(p_{O_2}) = -0.461 \cdot \log(p_{CO}) - 0.168$. Regime B shows mostly the CO covered surface at high pressures of CO and low pressures of oxygen. In this model, only the interaction term contributes to the SPTV model with a large positive coefficient. The α parameters surprisingly show a higher sensitivity towards CO, despite the surplus of CO being adsorbed on the surface. This might be a result of the overall higher CO desorption rates compared to O_2. However, as the SPTV models operate on the logarithm of both input and output values, it is difficult to interpret these parameters in a straight forward manner.

Table 3.5: Sensitivity parameters $\boldsymbol{\alpha}$ and LASSO coefficients $\boldsymbol{\theta}$ for the local regime models. Zero coefficients are represented by a dash for better readability.

Regime	α_{CO}	α_{O_2}	θ_{CO}	θ_{O_2}	$\theta_{CO \cdot O_2}$	θ^2_{CO}	$\theta^2_{O_2}$
A (transition)	-0.332	0.133	-	-	0.945	-	-0.162
B (CO covered)	2.333	-1.220	-	-	0.968	-	-
C (O covered)	4.088	-1.538	-	-0.191	-0.908	-	-

The surface phase diagram in regime C is dominated by adsorbed oxygen. The high sensitivity towards CO with an α of 4.088 results from only few adsorption sites being available for CO. Thus, the adsorption of CO is limiting the overall rate and increasing the CO partial pressure increases the probability of an adsorbed CO species. A negative sensitivity towards O_2 can also be understood as further adding more oxygen into an already oxygen covered system will poison the catalyst even stronger. In line with this interpretation is also the negative LASSO coefficient for the linear oxygen term in this model. Apart from that, the dominating LASSO term is again the interaction between the two reactants. For model C, however, the sign of this coefficient is negative, meaning that increasing both partial pressures in a similar degree will not result in a higher TOF.

The residual distributions for the three local regime models are given in Figure 3.19. Both, the models for regimes B and C show rather symmetric error distributions and small mean squared errors (RMSE) of 0.074 and 0.082 respectively. The errors in model A are significantly higher with an RMSE of 0.224. The extremely high maximum absolute error (MAE) for this model results from a single outlier as shown in the inset of Figure 3.19 a). This outlier is located close to the estimated transition between regime A and regime C. This proximity to the phase transition suggest a misclassification as the cause for this large error. In addition to that, the distribution for model A seems to be skewed towards positive values. This skewed distribution could be an indication that the kinetics in the transition region may require a more complex description than our SPTV model can provide. On the other hand, regime A covers only a small portion of the overall reaction condition space. The approximately uniform distribution of the QMC points result in regime A having the smallest training data set. Thus, the worse model performance compared to regimes B and C could also be a result of a less extensive training. A sampling procedure weighting all regimes equally despite those regimes covering differently sized portions of the factor space could therefore be a promising alternative to the QMC sampling.

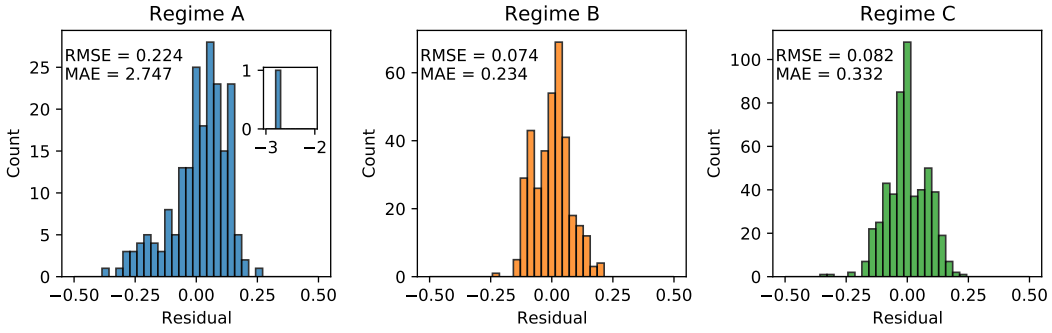


Figure 3.19: Residuals distributions for the three local regime models. Root mean squared errors (RMSE) as well as maximum absolute errors (MAE) are given in the figures.

Performance of the Global Model Combining the three local regime SPTV models with SVC predicted membership probabilities, we can set up the global surrogate model for the entire reaction condition space. Figure 3.20 gives an overview over the prediction performance of this global surrogate model. Comparing the predicted TOF over the entire partial pressure range in Figure 3.20 c) with the reference values in Figure 3.20 a), we can see that

the qualitative behavior of the MKM is well reproduced by our reduced model. The parity plot in Figure 3.20 b) shows that this agreement is even quantitatively accurate with an RMSE of 0.089.

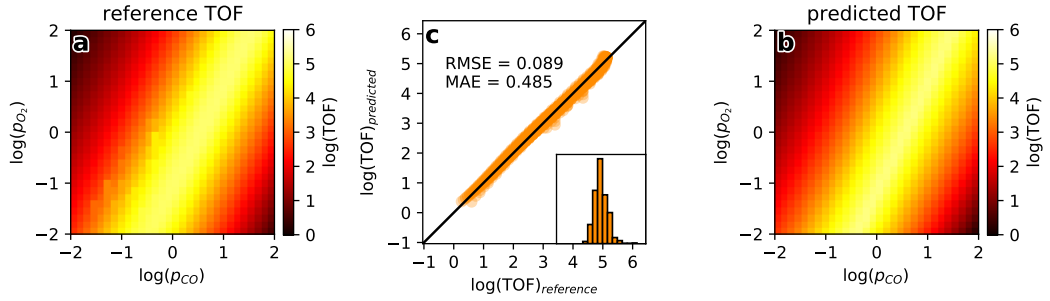


Figure 3.20: Prediction performance of the global surrogate model evaluated on a regular grid over the entire reaction condition domain. a) TOF calculated using the reference MKM. b) Parity plot between reference and predicted TOF values. The parity plot as well as the error values are determined on the test set depicted in the other subfigures. This test set is constructed as a regular 30 by 30 grid. c) TOF predicted by the global surrogate model.

To better assess the modeling accuracy of our approach, we construct regression models using several standard techniques based on the entire 1088 point data set. The results are summarized in Table 3.6. In these tests, our approach outperforms conventional polynomial regression models, which can not deal with the switching behavior at the regime boundaries. But also highly flexible methods like the decision tree based Random Forest Regression (RFR) and Gaussian Process Regression (GPR) show larger RMSE and MAE values compared to our coupled SPTV approach. The bad performance of these techniques can probably be attributed to their high demand in training data, especially in regions with high gradients as the phase transition in regime A. The only method outperforming our approach is Kernel Ridge Regression (KRR) with a radial basis function (rbf) kernel. While, this method is capable of accurately describing the entire partial pressure range, the resulting models lack interpretability. Our coupled SPTV model, on the other hand, does not only reveal the separation of the overall domain into local regimes, but also describes the local sensitivity of the reaction kinetics. Further, it should be mentioned that the good performance of the KRR(rbf) method relies on a high density of data points. For this reason, replacing the evenly distributed QMC points with a more sparse sampling of the experimental region could pose a problem for such methods.

Table 3.6: Comparison of the model root mean squared (RMSE) and maximum absolute errors (MAE) for the coupled local effective models with standard regression techniques: Linear and Polynomial Regression (Poly), Random Forest Regression (RFR), Gaussian Process Regression (GPR), Kernel Ridge Regression (KRR).

Method	RMSE	MAE
Linear Model	1.167	4.204
2nd Order Poly.	0.490	1.889
3rd Order Poly.	0.333	2.530
4th Order Poly.	0.220	0.841
RFR	0.175	1.630
GPR	0.259	6.471
KRR(poly)	0.329	2.431
KRR(rbf)	**0.047**	**0.424**
This Work	**0.089**	**0.485**

3.2.5 Extension of the CO Oxidation on RuO_2 to Variable Temperatures

Having investigated the isothermal CO oxidation process as a function of the partial pressures of CO and O_2, we now introduce temperature as a third factor. The overall procedure follows the same steps as previously for the two dimensional case. First, the overall parameter domain is sampled using QMC points from a Halton sequence. The reaction condition ranges are given in Table 3.7.

Table 3.7: Reaction condition ranges covered in the temperature dependent CO oxidation example.

Factor	-1	1
p_{CO} [bar]	$1{\cdot}10^{-2}$	$1{\cdot}10^{2}$
p_{O_2} [bar]	$1{\cdot}10^{-2}$	$1{\cdot}10^{2}$
Temperature [K]	550	650

An extensive data set of 512 QMC centers is used in this example, with a local 3 factor 3-Level FFD consisting of 27 points each. This results in a total data set size of 13824 points. For each LD, a linear SPTV model is constructed and the resulting model parameter space is analyzed by means of *Kmeans* clustering. As established in the previous section, temperature effects are considered in the form of $\exp(-\beta_M)$ values, such that the corresponding sensitivity parameter can be associated with an apparent barrier. The corresponding segmentation of the QMC centers is shown in Figure 3.21.

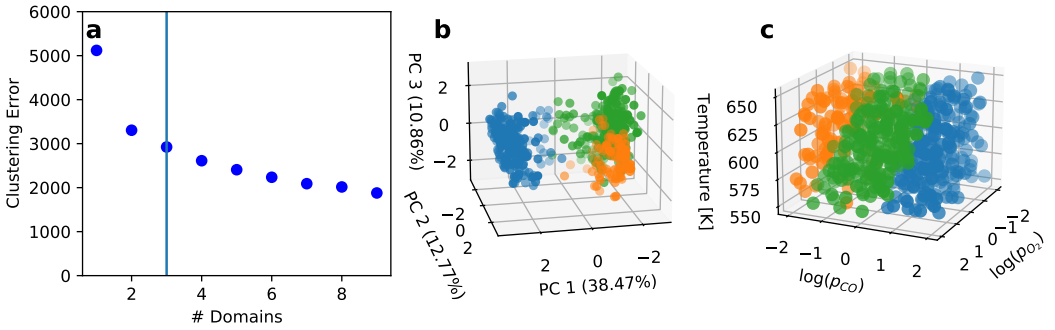


Figure 3.21: a) Scree-Plot for the temperature dependent CO oxidation on RuO_2. The vertical line indicates the selected number of domains $k = 3$. b) PCA representation of the model parameter space. The colors of the individual points indicate their clustering labels. The percentage of explained variance of the individual principal components is given in the axis labels. c) The center points represented in reaction condition space. The coloring, again, indicates the cluster assignment.

Again, a segmentation into three distinct domains is a reasonable choice. However, the PCA representation of the parameter space, shown in Figure 3.21 b), reveals significant overlap between the green and orange clusters. This might be the result of the three displayed principal components only accounting for roughly 62% of the variance in the data set. Still, the reaction condition representation of the corresponding segmentation of the QMC centers, shown in Figure 3.21 c), also shows considerable overlap with a much broader transition region (green cluster) compared to the isothermal case. Apart from that, a rather similar structure of the three clusters can be observed. Having identified the three regimes, the global surrogate model is constructed from three SPTV models with polynomial terms up to third order. The resulting model predictions have been evaluated on two 2D cuts through the reaction condition domain. The results are shown in Figure 3.22.

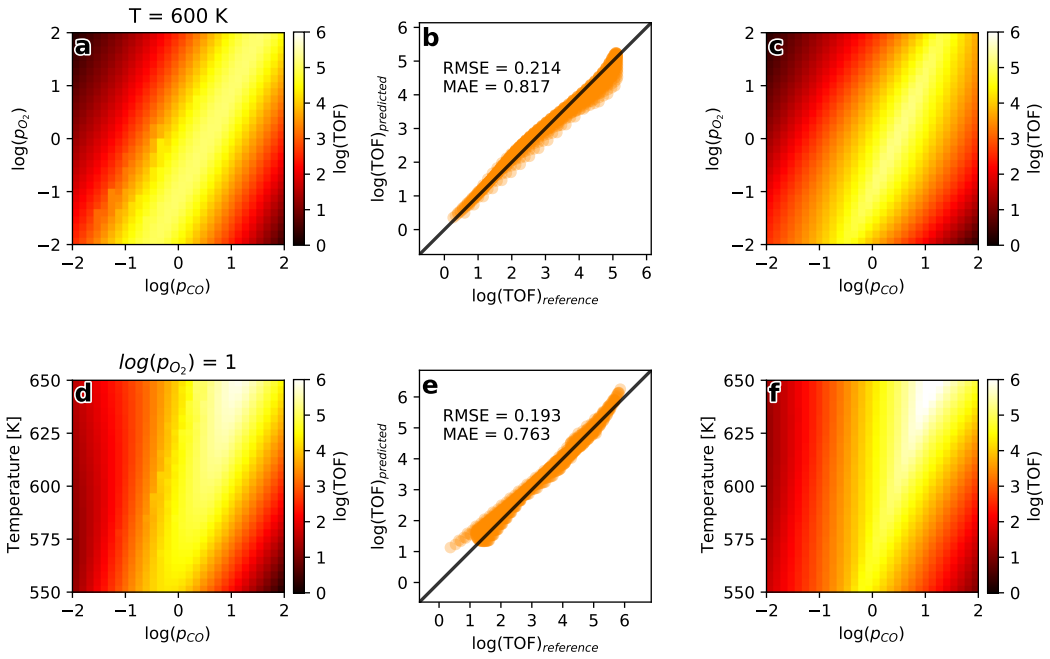


Figure 3.22: Prediction performance of the global surrogate model for the temperature dependent case evaluated on regular 2D grids cutting through the 3D reaction condition domain. a) TOF calculated using the reference MKM for a cut at constant T = 600 K. b) Parity plot between reference and predicted TOF values for a cut at constant T = 600 K. The parity plot as well as the error values are determined on the test set depicted in subfigures a) and c). This test set is constructed as a regular 30 by 30 grid. c) TOF predicted by the global surrogate model for a cut at constant T = 600 K. d) TOF calculated using the reference MKM for a cut at constant $\log(p_{O_2}) = 1$. e) Parity plot between reference and predicted TOF values for a cut at constant $\log(p_{O_2}) = 1$. The parity plot as well as the error values are determined on the test set depicted in subfigures d) and f). This test set is constructed as a regular 30 by 30 grid. f) TOF predicted by the global surrogate model for a cut at constant $\log(p_{O_2}) = 1$.

Figures 3.22 a), b) and c) show the results for a cut at constant T = 600 K through the three dimensional surrogate model. This cut corresponds to the same reaction conditions as previously investigated in the two dimensional case. With a RMSE value of 0.214, we can immediately see that the prediction quality is significantly worse compared to the temperature independent model. While the quantitative agreement with the MKM reference values is less pronounced, the qualitative behavior is well reproduced for this cut in the partial pressure plane. While the position of the regime boundary is accurate, we can see some additional curvature in Figure 3.22 c) compared to the reference. This might be an artifact of introducing temperature as the third factor. The rescaling of the temperature input, presented in the previous section, is a sensitive procedure and further assumes an Arrhenius like temperature dependence within the region of interest. If this assumption is not valid, systematic model deviations can occur. The second cut (Figures 3.22 d), e) and f)) covers various temperatures as well as CO partial pressures at constant oxygen pressure. The prediction quality of the global model is along this cut is comparable to the isothermal cut with a slightly lower RMSE value of 0.193. Looking at the qualitative agreement with the reference, we again observe accurrate reproduction of the MKM. Especially, the wedge shaped maximum TOF region in Figure 3.22 is nicely reproduced by the global surrogate model.

3.2.6 Outlook: Data Efficiency

We have shown that the combination of multiple local models can be used to describe chemical kinetics over larger domains of reaction conditions. The regime identification procedure we propose relies solely on the available data and does not require *a priori* knowledge of the reaction behavior or mechanistic details. For the CO oxidation reaction on RuO_2, the

SVM coupled surrogate models accurately reproduce the reaction kinetics of the mean field MKM at a fraction of the computational cost. In use cases like reactor simulations, where kinetic models are coupled to computational fluid dynamics, propagating a full MKM into steady state can become a computational bottleneck. While there is an initial investment necessary to set up the training data, in some scenarios, replacing the expensive MKM with a global surrogate could facilitate reactor simulations on significantly larger time and length scales. In general, however, reaction kinetics depend on way more than two factors and this leads to the main drawback of the naive QMC sampling of the global reaction condition domain. While QMC samples provide a highly efficient coverage of a multidimensional space, these samples are constructed for numerical integration of functions which show a certain degree of smoothness [10]. The behavior we see in reaction kinetics is exactly the opposite with regime boundaries leading to drastic changes in reactivity. Despite the fact that mean field MKMs show dramatically smoother behavior than e.g. kinetic Monte-Carlo simulations (which, on a microscopic level, are generally even more accurate for heterogeneous catalysis), we already run into problems looking at a 3D factor space. Introducing temperature as a third factor in section 3.2.5 results in a systematically lower performance of the global surrogate model. While the qualitative system behavior is still captured, a major drawback in this case is the huge amount of data required to set up this model. With a total of 13824 MKM simulation, the creation of the training data could already be infeasible for more complex reaction networks. This large data set, however, is necessary to locate the regime boundaries in the system with high accuracy. 512 QMC centers in three dimensions result in an approximately the same point density as the 64 centers used in the isothermal case. For a smooth function, such an dramatic increase should not be necessary. Locating a lower dimensional subspace, like the regime boundaries, on the other hand, becomes more and more difficult in higher dimensions. What makes the data efficiency even worse, is that most of the QMC centers are located within approximately smooth regimes, where the local behavior does not change much over a larger range of conditions. For these reasons, we require a refined sampling of the reaction condition ranges, to improve on the data efficiency of the approach and perform only those simulations necessary to on the one hand locate the regime boundaries and on the other hand to perform a polynomial approximation the local kinetic behavior.

3.3 Sequential Adaptive Designs for the Identification of Regime Boundaries

In this chapter an adaptive design approach for the efficient identification of kinetic regime boundaries is presented. The proposed algorithm builds on established optimal design theory [11] and introduces some necessary modifications to match the problem at hand. The first section gives an overview over the implementation of this adaptive design algorithm and how it differs from classical methods. This is followed by application test cases for various toy systems to assess the algorithm's performance. Finally, we apply our method to the previously studied MKM for the CO oxidation on RuO_2.

3.3.1 Introduction

The construction of a kinetic phase diagram for a chemical reaction system generally requires detailed mechanistic understanding. If, however, the reaction network is unknown and the only available information is the empirically observed dependence of the reaction rate on the applied conditions, we need an alternative way of determining such phase transitions. In the previous chapter, we could show that the behavior of local SPTV models can act as a fingerprint for the local effective reaction kinetics. The question now arises, how to most efficiently sample the space of reaction conditions to estimate the position of relevant phase transitions in a kinetic system. In order to fit a kinetic model over a range of reaction conditions, classical experimental designs are a valid option, if we can assume that the behavior within the desired range is smooth. Phase transitions mark the opposite of such smooth behavior and therefore, classical designs may not be appropriate for this class of problems. The

results of the previous sections have shown that fitting a local SPTV model over a smooth region of reaction conditions is fairly efficient in terms of required data points. Hence, within each kinetic phase (or regime) the required amount of data is low. What becomes the expensive part is localizing the position of phase transitions. To this end, an alternative to classical design of experiments can be optimal design theory, where the design matrix is numerically optimized under some model assumption in order to provide the best possible estimate of the model parameters [121, 122]. Therefore, the presence of kinetic phase transitions needs to be incorporated into the model assumptions for such an optimal design. As the position of these phase transitions is *a priori* unknown, a promising approach is sequential adaptive design [11]. In adaptive design theory, measurements are performed sequentially one after the other. By doing so, we can utilize all relevant information from the previous experiment to identify the most promising next sample point. This idea of sequential experimentation is similar to methods like active learning and bayesian optimization [123], which already found application in the context of catalysis research [124–128] and kinetic modeling [129, 130]. In these methods, the selection of design points is in general governed by a so called exploration-exploitation trade off. Exploration refers to regions, where the uncertainty of the current model estimate is high. On the other hand, a focus on exploitation would push design points towards regions where the current model estimates have some desired properties, e.g. the predicted TOF shows a maximum. The trade off between these two goals is defined by the acquisition function used in the optimization [131]. These techniques also rely on very flexible regression models like e.g. gaussian process regression [132] which can be demanding in terms of training set size. In contrast, sequential adaptive designs are constructed in a way, such that certain optimality criteria w.r.t to an assumed model class are fulfilled. For example, the error on the estimated model parameters should be minimized. Exploitation, in the sense of finding the optimum function value is not a primary focus of such design strategies. Of course, having determined a complex enough surrogate model, an optimization can still be performed in a subsequent step. In this work, we propose an sequential adaptive design algorithm based on the work by Fedorov [11, 54]. We assume a coupled multi-regime SPTV model to incorporate the phase transitions directly into the optimal design algorithm. By combining this classical theory with modern optimization and classification algorithms, we are able to construct optimal adaptive designs for reaction condition ranges which cover multiple effective regimes.

3.3.2 Modified Fedorov Algorithm for Regime Boundaries

We here describe a modified version of the sequential design algorithm for nonlinear parametrization based on Fedorov "Theory of Optimal Experiments" [11] (cf. section 2.3.4). We want to construct the optimal design for estimating the parameter of a global surrogate model for a reaction condition range, which may consist of multiple effective regimes. The optimization target, which dictates the position of the next measurement consists of two contributions, the efficiency function $\lambda(x)$ and the dispersion function $d(x)$. The main goal of this algorithm is to identify the regime boundaries as efficiently as possible. In order to achieve this, some modifications to the standard algorithm are introduced, which modify the estimated efficiency function for the next measurements.

Dispersion Function $d(x)$ In the optimal design framework, the dispersion function describes the estimated gain in information obtained by sampling and adding a point x to the design. This information content depends on the type of model we want to fit based on our design. For the functional form of the global model, we again use a set of coupled local effective SPTV models (equation 3.15) as in the previous section. While such polynomial models are linear in the coefficients (as the initial power transformation parameters are determined as part of the regression), we expect a nonlinear behavior in the boundary region, where local models of multiple regimes contribute to the overall global model. For nonlinear models the dispersion function $d(x)$ is calculated as given in equation 2.20. This equations contains the partial derivatives of the global model with respect to the model coefficients θ. As we now have a set of coupled local models, only the derivatives with respect to coefficients corresponding to the regime where x is located in, will contribute. Of course, in the boundary

region, contributions from all relevant local models are expected. In practice, these partial derivatives are evaluated as finite differences with small displacements in θ.

Efficiency Function $\lambda(x)$ The efficiency of an experiment is related to the information gain in performing this measurement. Generally, $\lambda(x)$ is proportional to the inverse of the expected error at a given point x. This means that the expected gain in information is larger, choosing a point x_{N+1} in a region where the predicted measurement error is small and the data is thus more reliable. In our setup, we additionally incorporate the position of the currently approximated regime boundaries. The efficiency function λ is constructed in the following way:

$$\lambda(x) = \frac{1}{\epsilon(x)} \prod_j \left[1 - \exp\left(\frac{-r_j^2}{2w^2} \right) \right] \quad \forall \quad r_j \leq r_{cutoff} \tag{3.16}$$

Here, $\epsilon(x)$ is the error model. The other product terms in $\lambda(x)$ are inverted gaussian terms with a minimum at $r_j = 0$ and standard deviation w. These terms do not depend directly on the values of the control variables x but on the distances r_j to the separating hyperplanes of the SVM. They correspond to the derivative of a hypothetical transition function between multiple domains. In practice, these terms act as a penalty for placing sampling points too close to the boundary region. In order to restrict these contributions to the transition region, a cutoff distance r_{cutoff} is introduced.

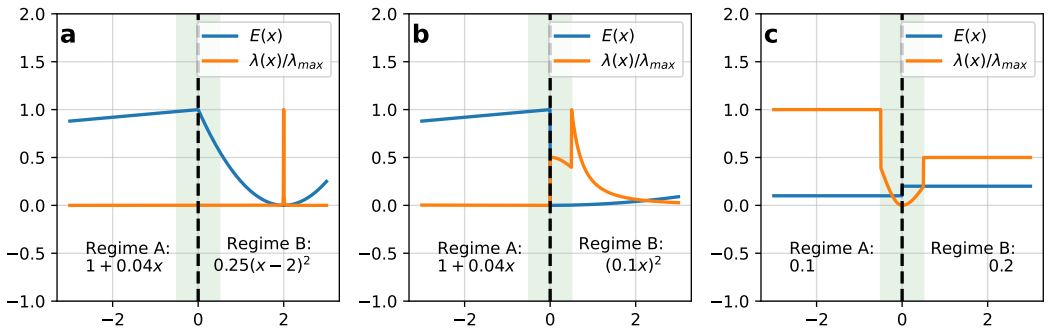


Figure 3.23: Behavior of the efficiency function $\lambda(x)$ for three different exemplary cases: a) The error model has a clear minimum value within regime B leading to a clear efficiency maximum at $x = 2$. The dip in efficiency around $x = 0$ is not resolved in this illustration due to scaling. b) The error function has a minimum within the regime boundary region. In this case, the repulsion term in $\lambda(x)$ leads to the maximum being located on the edge of the boundary region at $x = r_{cutoff}$. c) The error model is constant in both regimes. Here, also $\lambda(x)$ is constant, except for the boundary region, where the repulsion term leads to a low efficiency. The green shaded area corresponds to the cutoff distance r_{cutoff} from the estimated regime boundary ($x = 0$).

In later iterations of the adaptive algorithm, when the local models for the individual regimes are well determined, an additional biasing potential can be added to the efficiency function to more accurately determine the position of the regime boundary. For the functional form of this potential we can use an inverted form of the penalty in equation 3.23. This bias has its maximum directly at the regime boundary ($r_j = 0$) pushing the next design points towards the boundary. However, the width of this term is broadened compared to the penalty term, leading to an optimum of the resulting efficiency approximately at the cutoff distance from the boundary. By multiplying this term onto $\lambda(x)$, the effect of the penalty is not canceled out:

$$\lambda_{biased}(x) = \lambda(x) \prod_j \exp\left(\frac{-r_j^2}{4w^2} \right) \quad \forall \quad r_j \leq r_{cutoff} \tag{3.17}$$

Error Model The efficiency function $\lambda(x)$ relies on knowledge about the error of the model. For this reason, we simultaneously train the error model $\epsilon(\boldsymbol{x})$ while constructing our sequential design. While the absolute accuracy of the error model is not too important, the model should give a good indication of the relative errors over the entire reaction condition range. The model we employ here is a slightly modified linear regression model. Similar to the global kinetic model (eq. 3.15), also the error model $\hat{\epsilon}$ is composed of contributions from every regime coupled via membership probabilities obtained from the classification model:

$$\hat{\epsilon}(\boldsymbol{x}) = \sum_k p_k(\boldsymbol{x}) \epsilon_k(\boldsymbol{x}, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) \tag{3.18}$$

where the ϵ_k are the error models for the individual regimes. These models depend on the transformation parameters $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ as the linear regression is performed in terms of the power transformed variables in order to qualitatively capture nonlinear effects. One drawback of a linear model is that extrapolated values far from the training points may become negative. Negative values for the model error, of course, are not reasonable. To overcome this issue, we introduce an exponential transformation on the predicted errors, which ensures positive values. This transformation is scaled in a way that the resulting function does not deviate too much from a linear behavior, meaning that the linear term dominates the series expansion of the exponential within the region of interest. Again, the absolute value of the error is not relevant as long as the qualitative behavior of the error over the modeling range is captured. In addition to that, penalty terms are added to every data point proportional to the magnitude of its residual in the regression model. These penalties try to drive the adaptive design algorithm away from regions with an exceptionally large error that would lead to outliers in the linear regression. This may for example occur in the initial design steps, when the currently estimated regime boundary is still far away from the actual boundary. Measuring a point close to the real boundary will result in a large error for the respective local model and this is captured in the penalty term included in the error model. For a two dimensional problem depending on predictor variables x_1 and x_2, the local error model for a given regime may look as follows:

$$\begin{aligned} \epsilon_k(\boldsymbol{x}, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) &= e_1 \cdot \exp\left(e_2 \cdot \left(c_0 + c_1 \cdot x_1^{\alpha_1} + c_2 \cdot x_2^{\alpha_2}\right)^{\beta}\right) \prod_j P_j(\boldsymbol{x}) \\ P_j(\boldsymbol{x}) &= p_s \cdot |\rho_j| \cdot \exp\left(\frac{-1 \cdot ||x - x_j||_2^2}{p_w}\right) + 1 \end{aligned} \tag{3.19}$$

where $P_j(\boldsymbol{x})$ is a 2D Gaussian function with its amplitude scaled proportional to the residual ρ_j from the linear error model at training point x_j. p_s and p_w are hyperparameters of the penalty function. These Gaussian functions decay towards one, such that multiplying them onto the fitted error model does only influence the predicted values in the proximity of the training points. Typical values for the hyperparameters are $p_s = 0.1$ and $p_w = 0.04$. Here, especially the choice of p_w can be motivated by the desired sample density of the final design, as the maximum repulsion distance of the design points should be related to its support in design space. In practice, these parameters can be kept constant for different problems, as can the standard deviation of the Gaussian penalty terms. This is due to the fact that the adaptive design algorithm always operates on the scaled factors on the interval $[-1, +1]$. Thus, the absolute values of any factor do not influence the choice of these scaling parameters. A similar argument can be made for the choice of e_1 and e_2, the scaling parameters of the exponential transformation.

Optimization Problem Ultimately, the optimization problem which needs to be solved to locate the best subsequent point x_{N+1} is the same as equation 2.20. In order to solve this problem, we use a particle swarm optimizer. In this implementation the pyswarm package [133] is employed.

Batch Sampling In a real experimental study, it may often be useful in the adaptive design procedure not to add one single experiment at a time. Instead, a batch of N_b experiments may be required in order for the experimental workflow to run more efficiently. The problem of selecting a batch of experiments corresponds to finding the N_b best and most diverse points in terms of the adaptive design optimization function. We propose two different methods for approximating this without significantly increasing the computational cost of the optimization:

1. Batch Optimization: The first option is to extend the optimization problem from a N_f (number of factors) dimensional to a $N \cdot N_f$ dimensional problem by simultaneously optimizing the N_f coordinates of all N_b points contained in the batch. The diversity of these points is then emphasized by adding a repulsive potential between the batch points. Instead of N_b N_f dimensional optimizations, only one higher dimensional problem needs to be solved.

2. Biasing Potentials: The second option is to consecutively perform the optimizations for every of the N_b points. As soon as one optimum is identified, a biasing potential is added to the cost function at this location. This potential drives the optimization for the remaining points of the batch away from the already identified high dimensional design minima.

In this work, we have implemented option 1, the batch optimization. For the low dimensional problems illustrated here, the extended dimensionality $N_b \cdot N_f$ does not exceed the capabilities of the PSO for reasonable batch sizes. For example, a batch with 4 new experiments in a 2D kinetic experiment requires one 8D optimization to be performed. For the repulsive potential between the individual points we use a simple Coulombic $1/r^2$ term. The cost function for the batch optimization is

$$\begin{aligned} &\min_{\boldsymbol{x_b}} \sum_{n=1}^{N_b} -1 \cdot (\lambda(\boldsymbol{x_n}) d_N(\boldsymbol{x_n})) + \sum_{i \neq j}^{N_b} \frac{q^2}{r_{ij}^2} \\ x_b = [&\underbrace{x_{11}, x_{12}, ..., x_{1N_f}}_{\text{batch point 1}}, \underbrace{x_{21}, x_{22}, ..., x_{2N_f}}_{\text{batch point 2}}, ..., \underbrace{x_{N_b1}, x_{N_b2}, ..., x_{N_bN_f}}_{\text{batch point } N_b}]^T \end{aligned} \tag{3.20}$$

where r_{ij}^2 are the pairwise distances between the individual batch points and $\boldsymbol{x_b}$ is the $N_b \cdot N_f$ dimensional vector of the batch coordinates. The coulombic repulsion between the batch samples is governed by the pseudo charge q assigned to each point. The overall optimization is very sensitive to this hyperparameter, which determines how diverse the identified minima will be. To loosen the requirement of identifying the optimal q value, all batch minima are locally relaxed without the repulsive term after the combined optimization. This way, individual minima should be discovered, as long as all solutions ended up in the correct basin of the cost function during the initial step. The dependence on q is illustrated in the example in Figure 3.24.

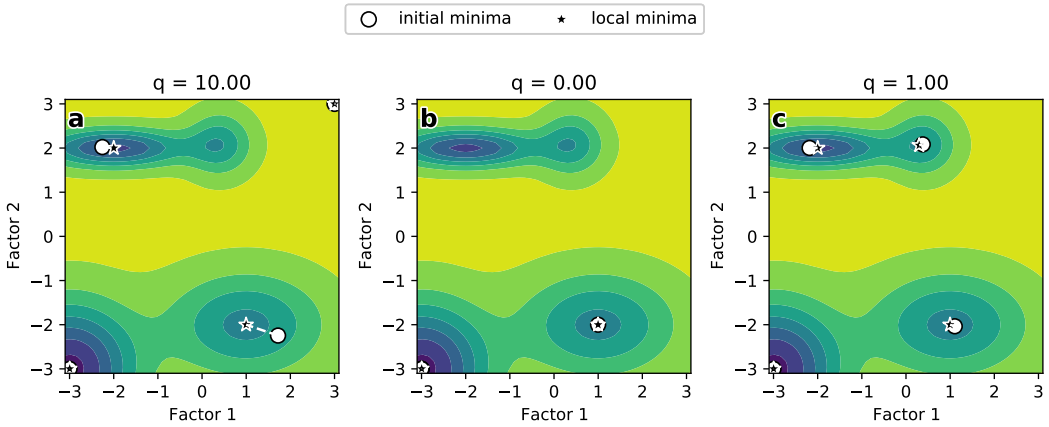


Figure 3.24: Batch optimization for a 2D function with four minima. The initial minima obtained in the batch optimization are marked by white circles and the locally relaxed solutions are indicated by black stars. The corresponding pairs are connected by a white dashed line. a) Optimization result for $q = 10$, b) for $q = 0$ and c) for $q = 1$.

For a high pseudo charge (Figure 3.24 a), the repulsion between the individual solutions is large. For this reason, the two minima located closest to each other are not resolved, but one of the sample points is pushed towards the overall boundary. The other extreme is an optimization with zero charge (Figure 3.24 b). Here, we can see that only two of the four minima are identified, as the individual samples feel no repulsion among each other. This way, all solutions tend to converge towards the lowest minima. Figure 3.24 c) shows the results for $q = 1$. Using a moderate repulsive strength, the algorithm is successful in identifying all four minima of the function. In general, this choice of q may not be transferable two arbitrary optimization targets, as the optimal hyperparameter value depends on both the distance between and the depth of the minima.

Design Initialization The general Fedorov algorithm relies on a nonsingular design for the initial set of points. In our setup, a key requirement for the overall design to be nonsingular is that the subsets of design points located in the individual regimes are sufficient to determine the regression coefficients of the respective local model. As, however, for the initial design, the regime boundaries are not known, a priori defining a nonsingular design can be challenging. To avoid this issue, we introduce an additional initialization step before relying on the Fedorov algorithm to select the next points. The overall design is initialized with a user selected standard design, e.g. a Placket-Burman type screening design. Now, all local models are determined and initial regime boundaries are obtained. At this stage, the initialization algorithm checks, if each subset contains enough data points to fit a linear local regime model. If this requirement is not met by every regime, additional sample points are selected in this regime. These points are determined by employing a reduced adaptive design algorithm. In the absence of both, an estimate for the global model as well as the error model, we have no information about the efficiency nor the nonlinear dispersion function. Therefore, we use the linear D-optimality criterion $\max |X^T X|$ separately for every regime, where X is the design matrix for a linear regression model. By maximizing this determinant, the distance between design points will be maximized within each regime. Such ideas are closely related to techniques like farthest point sampling, where one samples as diverse sets of points as possible [134, 135]. This reduced algorithm can also be described as a Fedorov algorithm for linear parametrization with a constant efficiency function $\lambda = const$. This procedure is repeated until each regime provides sufficient samples to fit an initial global model. During these optimization steps of the individual regime designs, of course, the penalty with respect to the distance of the samples to all regime boundaries is considered in the same way as later in the full Fedorov optimization. A 2D example of the here described design initialization is given in Figure 3.25. In this example, the initial global design consists of only five points. Analysis of the initial local models gives rise to a segmentation of the

parameter space into three regimes. As each regime is supposed to be described by a linear local model, at least three sampling points per regime are required to estimate both slopes and the intercept of the model. In this situation the initialization procedure is triggered for each regime and additional points are added to the design.

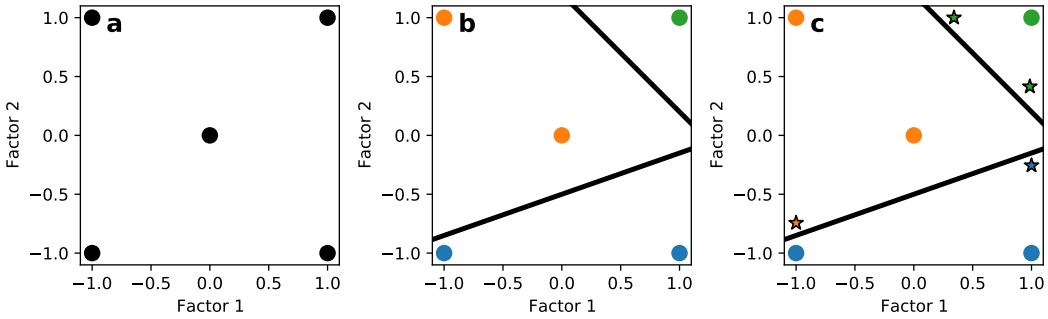


Figure 3.25: Example for a design initialization step in 2D. a) The global design has been initialized with a two factor two level full factorial design sampling all corners of the square parameter space. Additionally, the central point has been added. b) Segmentation of the parameter space according to the initial local models. c) Extended design matrix after the design initialization step. The additional points added in the initialization step are indicated with stars.

All newly added points in Figure 3.25 are located close to the vertices of the individual regime facets. This way, the area covered by the design is maximized. The residual deviation from the vertices is intentional and caused by the repulsive penalty close to the regime boundaries.

Piecewise Approximation of Nonlinear Boundaries Assuming linearity of the regime boundaries is, of course, a simplification in the absence of more information. Like shown in the previous chapter, this approximation often works well. However, to deal with the more general case of nonlinear boundaries, we need to expand the classification procedure used in the adaptive design algorithm. Using a nonlinear kernel for the SVC model provides the required flexibility to identify nonlinear boundaries, but the calculation of the efficiency function λ relies on a well defined distance metric orthogonal to the regime boundary. Using nonlinear kernel functions, like radial basis functions (rbf) or polynomials, it is difficult to determine these distances, as the decision boundary of the SVC is no longer a simple function of the design variables. For this reason, we resort to use a piecewise linear approximation to the nonlinear regime boundary to keep the advantages of simple linear functions like e.g. the data efficiency. To this end, we start with fitting the classification model using a very flexible rbf kernel function. Based on this model, we can predict the score of any given point x. For the linear SVC case, these scores were directly proportional to the distance from the decision boundary. While this property is not strictly true for the nonlinear case, the magnitude of the SVC score is still related to this distance. By evaluating the SVC score on a dense grid over the entire range of design variables, and setting a threshold for this score, we can obtain a point cloud, located in the vicinity of the nonlinear boundary. An example is given in Figure 3.26 a), where the data points exactly follow the parabolic shape of the regime boundary. In order to segment these points into multiple parts along the boundary, we can project them onto their main principal component and partition them according to a set of percentiles like shown in the inset of Figure 3.26 a). We end up with a piecewise linear model, by taking always the end points of each segment and fitting a linear model in between. In case the outer most points do not fall onto the overall boundaries of the design space, we can determine the closest edge and add an additional point there. The result is shown in Figure 3.26 b).

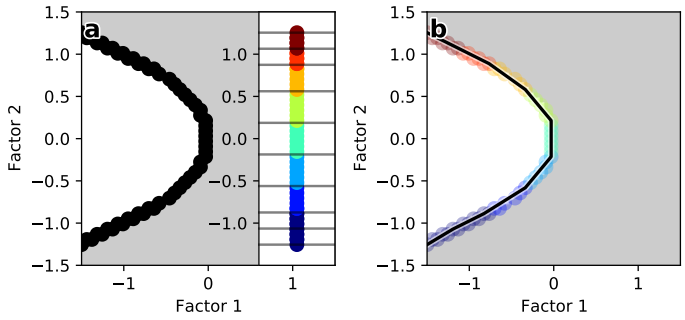


Figure 3.26: Illustration of the piecewise approximation of nonlinear boundaries. a) 2D design space with two regimes separated by a parabola shaped boundary. The black data points have a SVC score < 2. The inset shows the projection of these points on the main principal component with the colors indicating the segmentation according to several percentiles. b) The representation of the segmented data set in the original 2D space. The line shows the piecewise approximation to the decision boundary.

Overview Figure 3.27 shows a schematic of the adaptive design algorithm for the identification of regime boundaries. Starting with an initial set of points, first it needs to be determined, whether this initial global design results in nonsingular partial regime designs. If this is not the case, the design initialization procedure adds points to the global design based on the adaptive design criterion for linear parametrization. During this phase, the efficiency function λ does not contribute to the optimization, as it depends on the error model which is not yet well determined due to a lack of data. After each added point, the regime assignment gets updated until every identified regime contains sufficiently many data points. As soon as this criterion is met, the algorithm switches into the main optimization loop. First, the individual local regime models are constructed and combined into a global model. Next, a global error model is obtained in a similar fashion. The error values for every design point are based on the residuals of the local model used to obtain the local feature vector. Based on these two models we can determine the cost function for the main Fedorov algorithm. The dispersion contribution to the cost function depends on the global model's gradients w.r.t. the regression coefficients, while the error model directly enters the efficiency function λ. Depending on whether a refinement of the exact regime boundary position is desired or not, additionally the confining bias potential may be added to λ. Using a particle swarm optimizer, we determine the position of the next sample. Having obtained the coordinates of the next point, the algorithm checks, if the local design around this point would overlap with any existing design points. If this is the case, instead of adding a new point to the design, the closest already known point is reevaluated to obtain a better estimate of the corresponding measurement error. If the optimized sample is actually new, the corresponding point is added to the overall design. Finally, based on the updated design, the new position of the regime boundaries are estimated. This loop continues until some convergence criterion is met. The choice of the convergence criterion depends on the specific problem and the goal of the adaptive design procedure. If the main goal is to locate the regime boundaries, terminating the algorithm as soon as the position of the estimated boundary does not change beyond a certain threshold seems reasonable. However, at this point, the design might not yet be sufficient to accurately model the system behavior within each regime. Therefore, if the design is supposed to allow for an accurate modeling of the system, rather a criterion based on the dispersion contribution to the optimization cost function could be useful. In practice, however, most often the maximum number of experiments will probably be the limiting factor.

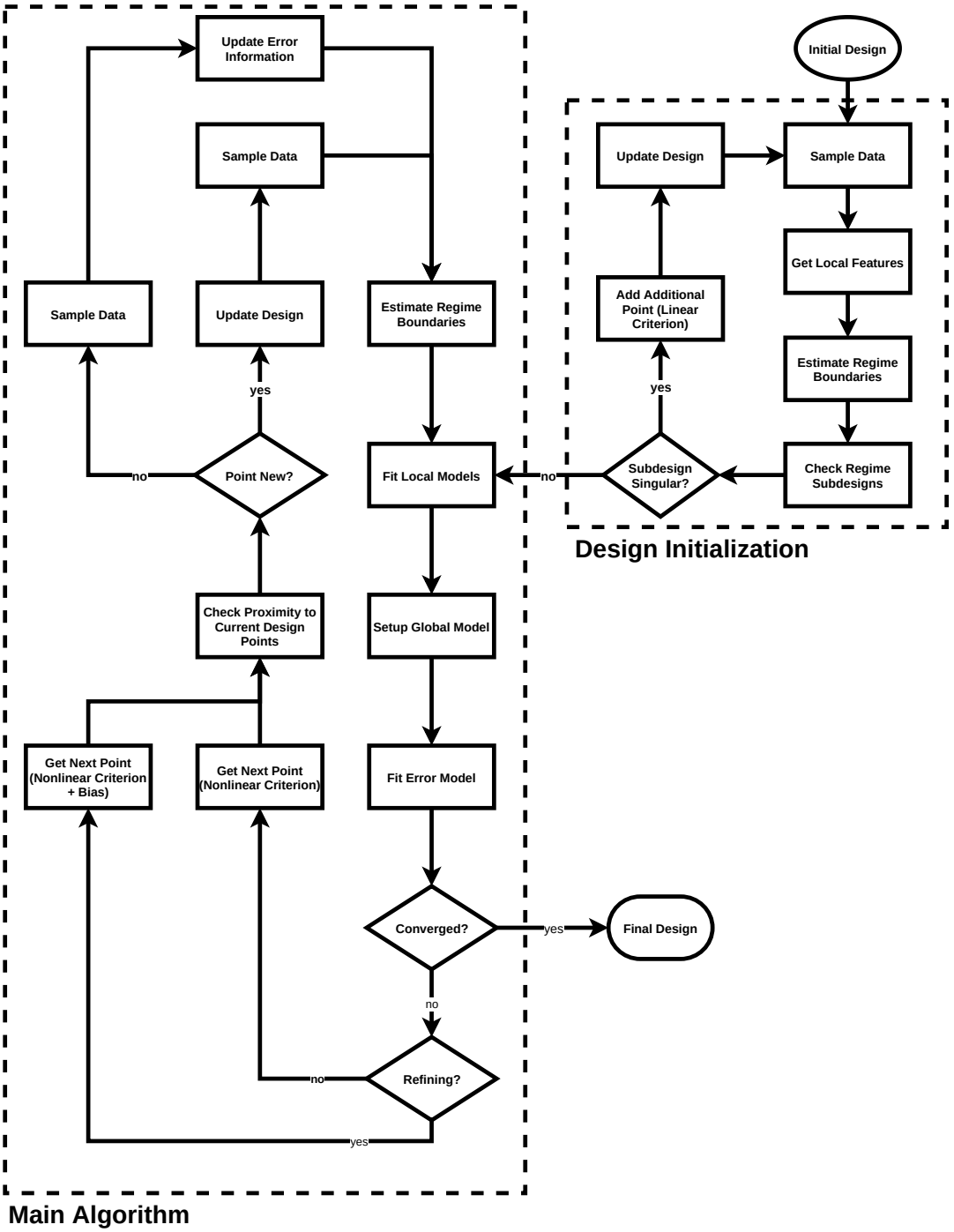


Figure 3.27: Schematic of the modified Fedorov algorithm for regime identification.

3.3.3 Algorithm Behavior for Several Toy Problems

To assess the performance of the proposed design algorithm, we constructed characteristic toy models for which the exact position of the regime boundary is known. These toy models consist of simple analytical functions within each regime coupled via a smooth transition function at the boundary. Additionally, statistical noise is added to the sampled data, both in terms of uncertainty on the input variables (x) and the response of the function. To model the behavior of real phase transitions more realistically, the standard deviation of the noise term on the response values is nonconstant over the range of input variables with its maximum at the regime boundary. For the purpose of illustrating the algorithm behavior, the examples

shown here, are 2D functions in variables x_1 and x_2. The analytical expressions as well as the final design matrices for these examples are given in the appendix (cf. A.3 and A.4).

Regime Boundaries in 2D: Constant Function

The simplest toy model of this type is a constant function which takes on different values depending on the respective regime. Further, the boundary between these regimes is linear, enabling the adaptive design algorithm to in theory find the exact solution in the limit of a large set of samples. Figure 3.28 shows the behavior of three piecewise constant toy systems, which possess different linear regime boundaries. The first model corresponds to a two regime system with one transition along a linear function in x_1 (Figure 3.28 a), while the other two toy systems represent two possible scenarios of a three regime problem.

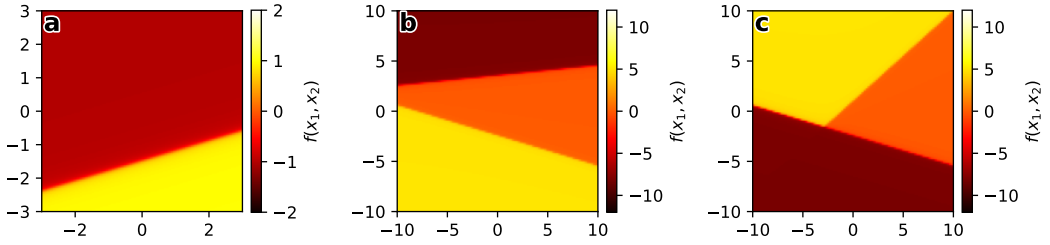


Figure 3.28: Toy models with linear boundaries and constant function values within each regime. While the "real" input parameter ranges for the individual models are different, the coordinates x_1 and x_2 are scaled to the same design range [-1,1] throughout the adaptive design process. a) Model with two regimes and one boundary. b) Model with three regimes and two transition boundaries. c) Model with three regimes and three phase transitions creating a triple point.

The first test of the proposed algorithm is conducted for the two regime system in 3.28 a). Without any knowledge on the system at hand, standard experimental plans are usually a reasonable choice for an initial design. For example, depending on the amount of experimental measurements feasible, we can initialize the adaptive design algorithm with a 2-Level full factorial design to investigate the effective model behavior at the extreme points of the design space. This way, the presence of regime boundaries separating the entire space, should directly be identified. In a 2D case, meaning two experimental factors, the 2-Level FFD consists of the four corners of the design space:

$$D_{initial} = \begin{bmatrix} -1 & -1 \\ +1 & -1 \\ -1 & +1 \\ +1 & +1 \end{bmatrix}. \tag{3.21}$$

As established in the previous chapter, the classification of the design points is performed based on local features. These features are based on the behavior of local models fitted to local experimental designs with small displacements around the central global design point. For the local designs, we again, use 2-Level FFDs with the addition of the central point $[0, 0]$,

which coincides with the global design point

$$D_{local} = \begin{bmatrix} -1 & -1 \\ +1 & -1 \\ 0 & 0 \\ -1 & +1 \\ +1 & +1 \end{bmatrix}. \tag{3.22}$$

The initial as well as well as the local designs used for the 2D to systems are schematically illustrated in Figure 3.29.

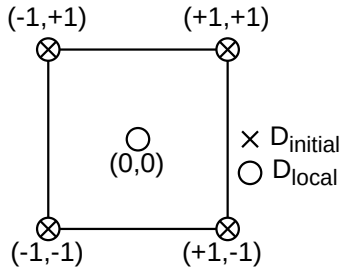


Figure 3.29: Schematic illustration of the initial and local 2 factor designs.

Investigating such simplified toy models, instead of fitting a SPTV model and clustering according to the LASSO coefficients and transformation parameters, locally constant models are used. Hence, the local models are solely distinguished by one parameter, the constant function value. After the initial regime identification, the design initialization step is performed in order to create locally nonsingular designs in every regime. As previously discussed, this is a requirement for setting up the local (per regime) linear regression models. While a linear model would not be necessary for these piecewise constant toy systems, the error model which enters the efficiency function λ still needs to be determined. Figure 3.30 gives an overview over the adaptive design for the toy system in Figure 3.28 a) after this initialization step.

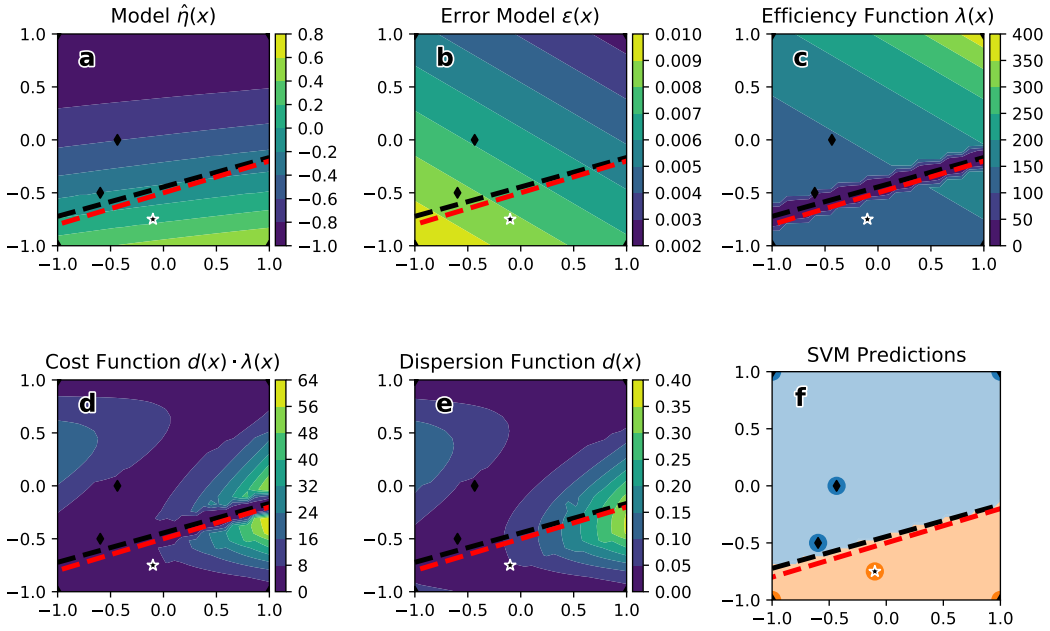


Figure 3.30: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 a). All functions are evaluated based on the design after the initialization step. a) The prediction of the current global model approximation. b) The prediction of the error model. c) The efficiency function for the optimization. d) The cost function for the optimization. e) The dispersion function based on the current global model. f) The clustering of the current data set and the resulting SVM predictions. The black dashed line corresponds to the current estimate for the regime boundary based on the SVM while the red line indicates the position of the known reference boundary. The star indicates the latest added point of the design.

As we can see, additional points have been added to the initial design $D_{initial}$. A linear model in two variables requires at least three data points for the estimation of the regression coefficients. However, the blue regime in Figure 3.30 f) contains four points already after the initialization step. This is due to the fact, that the regime assignments get updated after every added data point. Hence, it took the design initialization two iterations to set up this locally well determined design. The efficiency function is shown in Figure 3.30 c). By definition, λ behaves like the inverse of the error model ϵ. In addition to that, there is the penalty term in the proximity of the estimated regime boundary. The dispersion function in Figure 3.30 e) has a clear maximum around [1.0,-0.5] which tries to maximize the distance to the already known sample points. Combining the dispersion and efficiency functions, we end up with the cost function of the adaptive design algorithm (Figure 3.30 d)). The cost function now also shows the penalty around the boundary line, which pushes the optimum of the dispersion function away from the boundary. Figure 3.31 shows how the individual contributions are updated after sampling the next data point. We can see that the position of the estimated regime boundary is slightly shifted and now coincides well with the known reference (red dashed line). As a result of the additional data point and the shifted regime boundary, all contributions to the cost function also undergo some change. The most drastic difference can be observed for the dispersion function in Figure 3.31 e). The newly added data point reduced the dispersion contribution around its position giving rise to a new optimum around [-1.0,0.25] for the subsequent iteration. In general, the position of the regime boundary has been accurately determined after only one iteration of the adaptive design algorithm plus the initialization step with a final design consisting of only eight data points.

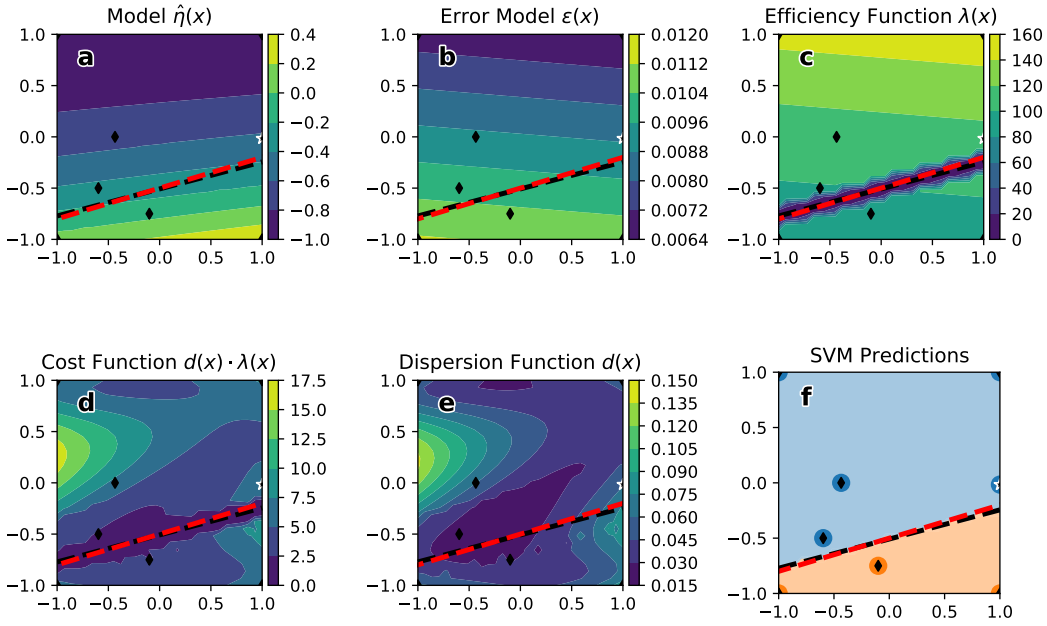


Figure 3.31: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 a). All functions are evaluated after one iteration. Subplots are the same as for Figure 3.30.

The toy system in Figure 3.28 b) shows a slightly more complex behavior. Now, three distinct regimes are present in the design space and, hence, two decision boundaries need to be determined. Again, we choose a 2-Level FFD for the initial design. As for the previous example, the first design points are already added during the design initialization. A linear model in three distinct regimes would require at least nine total design points. The initialized design shown in Figure 3.32 consists of 10 points, requiring only one additional point compared to the minimum number of nine for the initial global response model. The algorithm accurately identifies the three regimes, despite none of the initial 2-Level FFD points being located in the third regime (green area in Figure 3.32 f)). As the SVC classifier determines pairwise decision boundaries for each combination of classes, there are three estimated regime boundaries (black dashed lines) present in this three regime example. Following the adaptive design procedure for four iterations, both regime boundaries are well determined as shown in Figure 3.33.

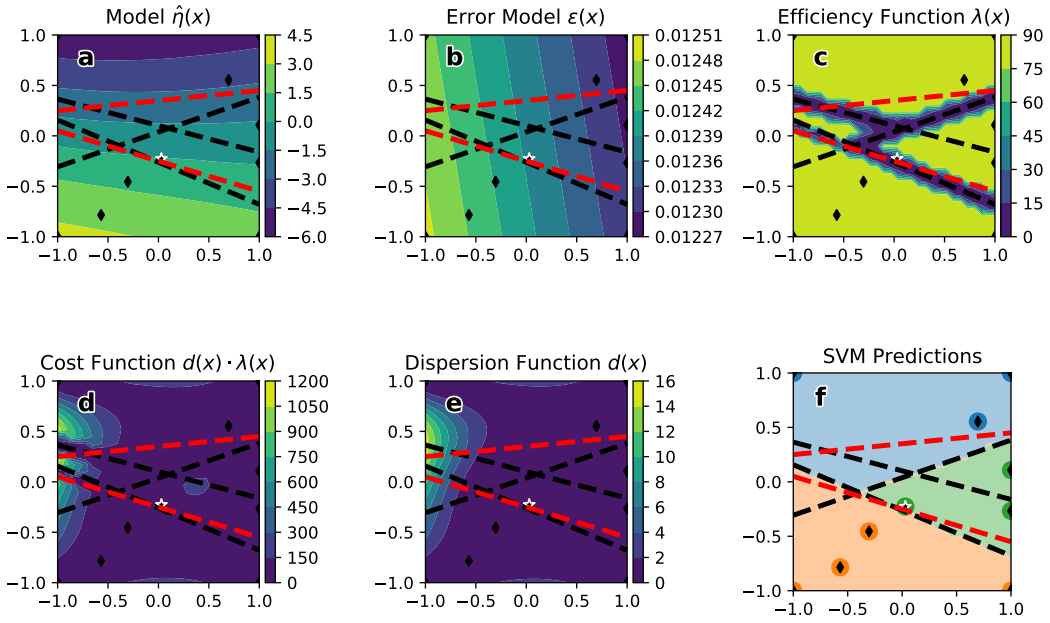


Figure 3.32: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 b). All functions are evaluated after the design initialization. Subplots are the same as for Figure 3.30. Despite only two regime boundaries (red dashed lines), there are three SVC decision boundaries as the classification model determines pairwise two class boundaries.

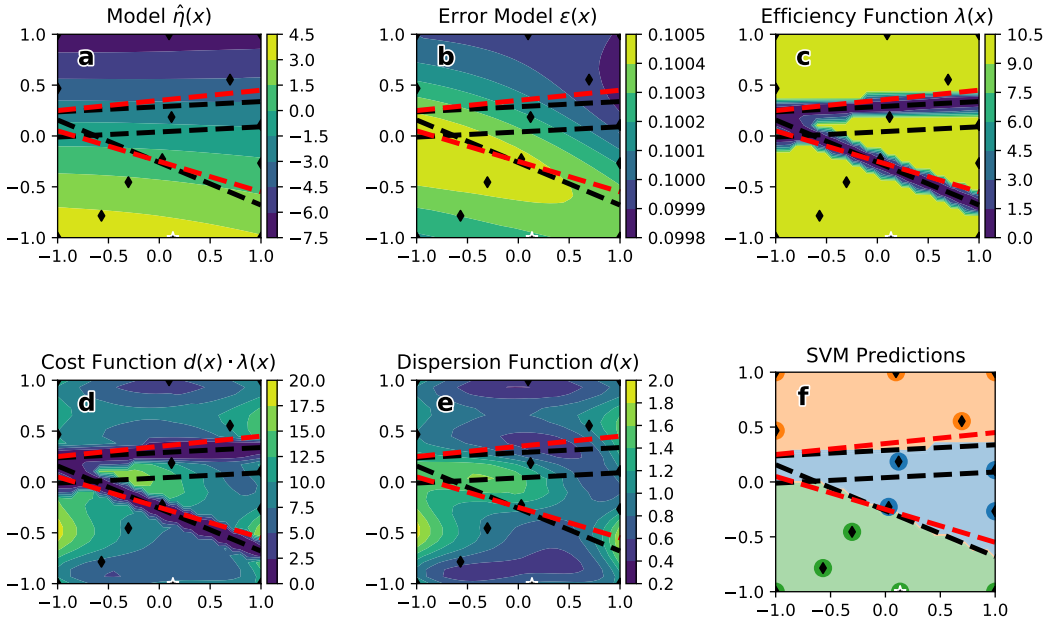


Figure 3.33: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 b). All functions are evaluated after four iterations. Subplots are the same as for Figure 3.30. Despite only two regime boundaries (red dashed lines), there are three SVC decision boundaries as the classification model determines pairwise two class boundaries.

The third piecewise constant toy system exemplifies a second possible three regime scenario where the two regime boundaries intersect and form a triple point in the phase diagram. For this toy system (Figure 3.28 c)), the design initialization gave similar results as for the previous example without the boundary intersection as can be seen in Figure 3.34. It took the algorithm six iterations to accurately estimate the position of the regime boundaries and the triple point of the system.

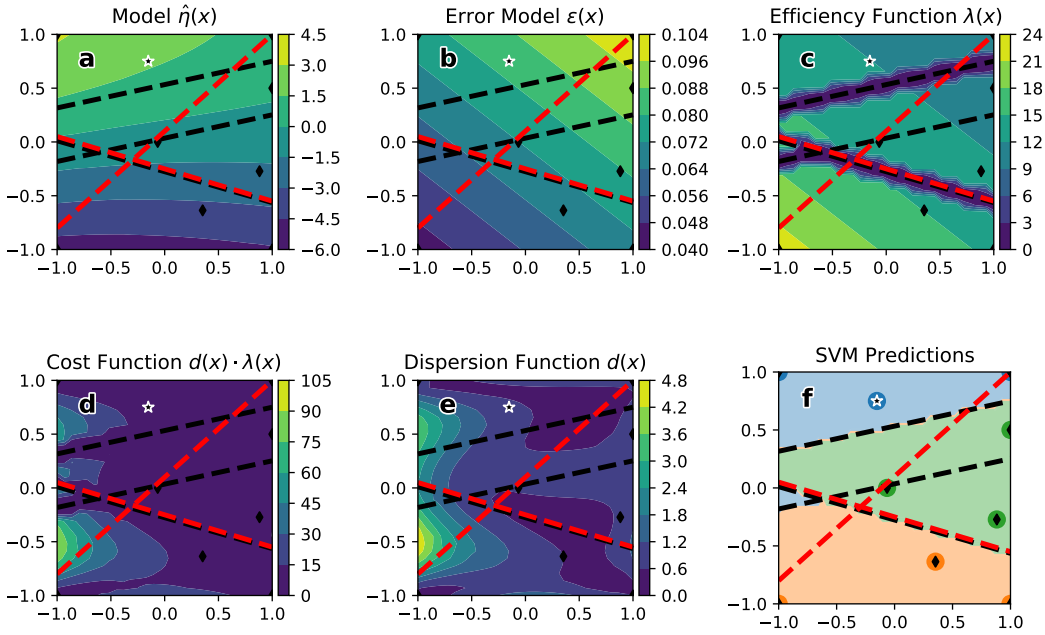


Figure 3.34: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 c). All functions are evaluated after the design initialization. Subplots are the same as for Figure 3.30. Despite only two regime boundaries (red dashed lines), there are three SVC decision boundaries as the classification model determines pairwise two class boundaries.

In this case, starting with the 5th iteration, the confining potential has been activated to more efficiently sample the proximity of the regime edges. Looking at the efficiency function in Figure 3.35 c), the effect of this confining potential becomes obvious. The estimated efficiency of the next measurements becomes maximal close (but not directly on top of) the estimated boundaries. These potential functions depend only on the distance to the respective SVC decision boundaries and are evaluated for each of them. Hence, both the repulsive and the confining potential terms multiply close to intersections of multiple boundaries. This emphasizes the sampling of design points close to the triple point of the phase diagram. In total, it took 15 designs points to get a reasonable approximation of the phase diagram of this toy system.

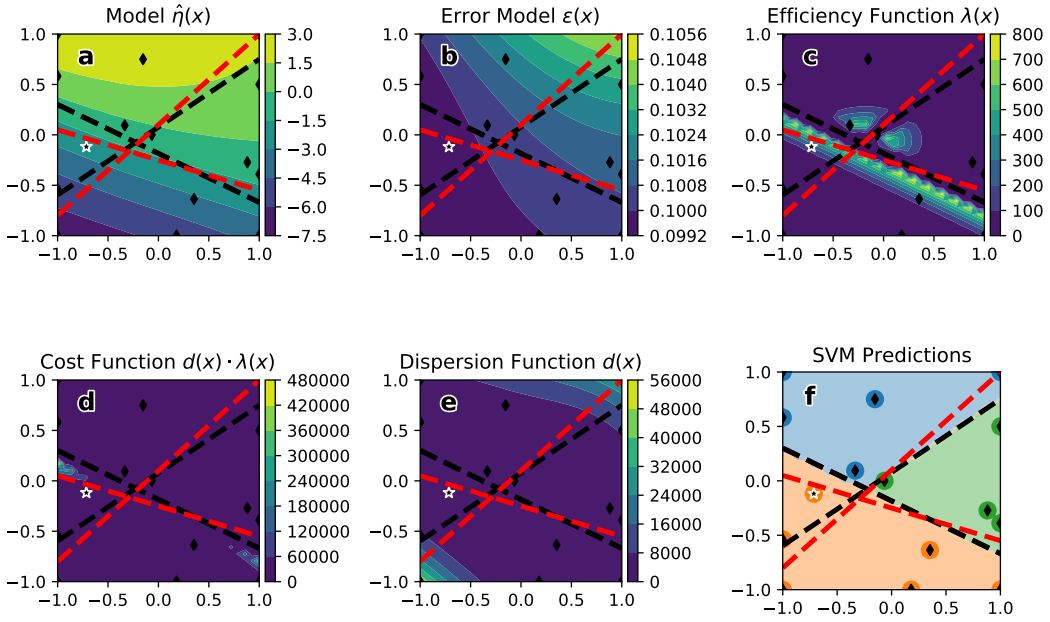


Figure 3.35: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 c). All functions are evaluated after six iterations. Subplots are the same as for Figure 3.30. Despite only two regime boundaries (red dashed lines), there are three SVC decision boundaries as the classification model determines pairwise two class boundaries.

Regime Boundaries in 2D: Nonconstant Function

In a second step, we investigated the regime identification capabilities of the algorithm for toy systems which show nonconstant behavior within each regime. Also for these examples, we did not use SPTV models for the local representation of the model function, as the underlying dependencies are less complex. The model representation used to obtain the set of local features for every global design point, is a LASSO regularized linear model. Due to the feature selection properties of the LASSO by clustering according to the coefficients of such a model, we can distinguish between local models depending on either x_1 or x_2 in a straight forward manner. The regime wide local models contributing to the overall global model, are then either linear or second order LASSO models depending on the amount of data points available per regime. Figure 3.36 shows the behavior of the two nonconstant toy systems. Both systems show a dramatically different functional dependence in the two regimes, with one regime depending only on x_1 and the other only on x_2. The major difference between the two systems is the shape of the regime boundary, which is linear in Figure 3.36 a) and follows a parabolic function for the system in Figure 3.36 b). For the initial and local designs, the same point sets have been used as for the piecewise constant toy systems.

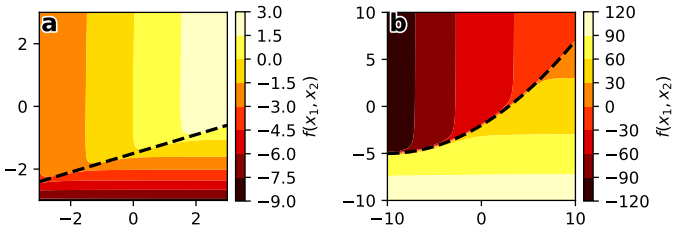


Figure 3.36: Toy models with nonconstant function values within each regime. a) Model with two regimes and one linear boundary. b) Model with two regimes and one nonlinear transition boundary.

Figure 3.37 gives an overview over the adaptive design for the toy system with the linear boundary. We can see that during the initialization step two new design points have been added meeting the minimal requirement of six data points for a 2D linear model in two regimes. While the assignment of each design point to its respective class is correct, the position of the initially estimated regime boundary is vastly different from the known reference. This is simply due to a lack of available data. The cost function in Figure 3.37 d) is dominated by the dispersion contribution which pushes the next design point towards the least sampled area of the design space.

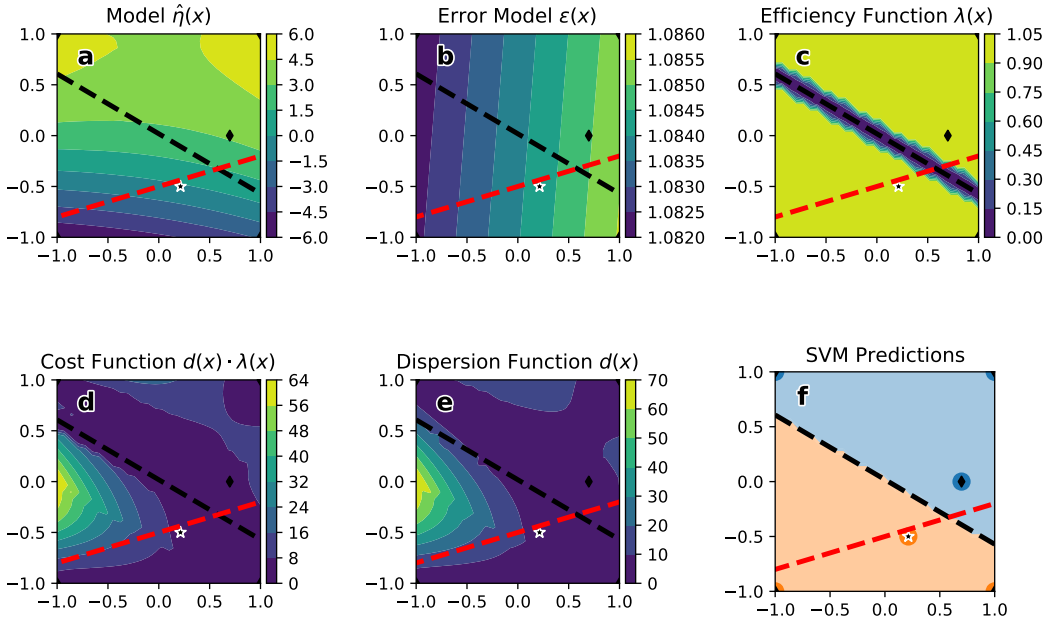


Figure 3.37: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.36 a). All functions are evaluated after the design initialization. Subplots are the same as for Figure 3.30.

After four iterations, the adaptive design algorithm has accurately determined the transition line between the two regimes. The efficiency function in Figure 3.38 shows that the confining potential has been activated and thus the next sample points would be located close to the regime boundary. The final design consists of a total of 10 samples, which mainly lie near the edge of the design space. As the functional depence within each regime is simple, the amount of points required for the sampling of the area of each regime is small. As soon as the

regime wide models are well determined, sampling more points on the edge of the design space can provide more information on the exact position of the boundary region. Initially, this trade off between refining the local regime models and localizing the phase transition is automatically regulated by the interplay of the dispersion and efficiency functions. At the point when sufficient data points are located in each regime, however, the dispersion contribution can become less and less pronounced. At this point, it can be beneficial to activate the confining potential to push the design points towards the decision boundary. The location of the cost function maxima along this boundary are then still dictated by the residual dispersion contribution.

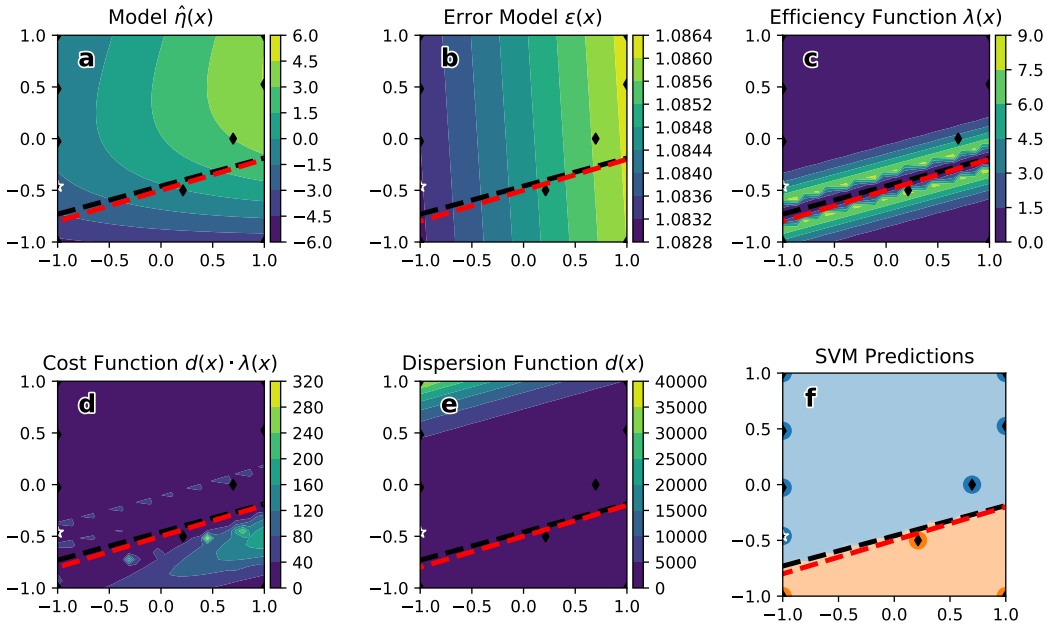


Figure 3.38: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.36 a). All functions are evaluated after four iteration. Subplots are the same as for Figure 3.30.

The second toy system in Figure 3.36 b) shows a nonlinear regime boundary. The purpose of this toy system, is to test the piecewise linear approximation of nonlinear SVC decision boundaries. For the design initialization procedure, we started with a linear kernel for the SVC resulting in the design illustrated in Figure 3.39. Similar as for the linear regime boundary, the classification of the initial data points is correct, but the estimated boundary deviates dramatically from the reference due to lack of data. Starting with the first adaptive design iteration, we use a radial basis function kernel for the SVC and apply the piecewise linear approximation procedure to the decision boundary. Whether to use a linear or nonlinear kernel in the classification model is currently a user supplied input. A possible solution, however, might be to define a error threshold for the linear SVC, which if exceeded triggers the usage of a nonlinear kernel. The resulting nonlinear classification model is shown in Figure 3.40.

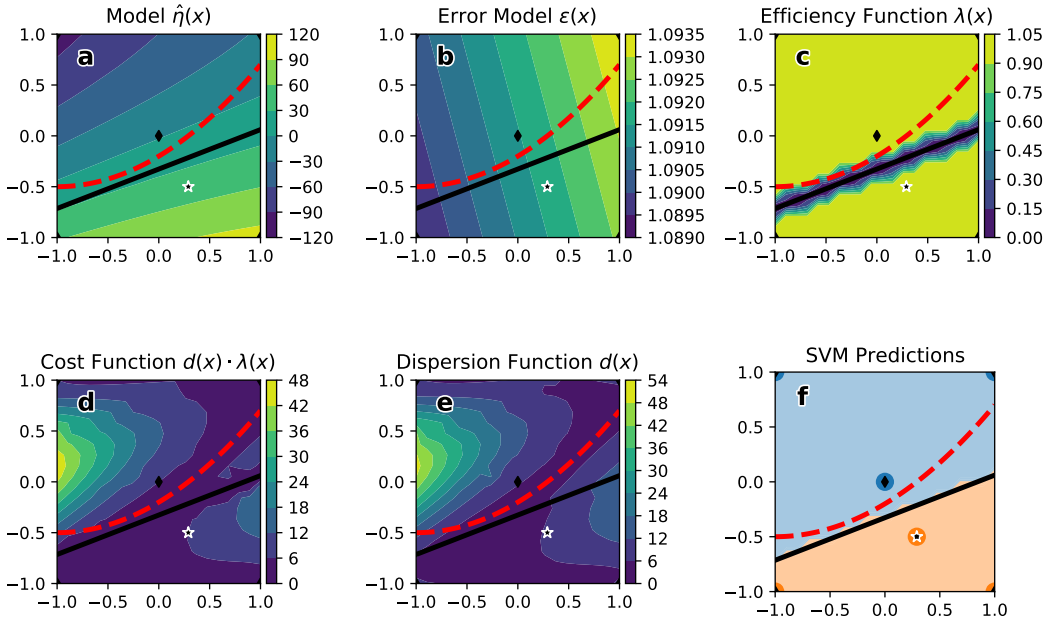


Figure 3.39: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 b). All functions are evaluated after the design initialization. Subplots are the same as for Figure 3.30.

After three iterations the algorithm gives a reasonable representation of the parabolic shape of the regime boundary. At this point, the partial regime designs consist of four sample points each. Therefore, the regression coefficients for linear response models in both regimes are well determined. This is reflect in the low values of the dispersion function in Figure 3.40 e). Further, all maxima are close to the already existing designs points in the corners and on the edges of the design space. While the number of data points is sufficient to determine the global response model, the distribution of the points over the design space is not optimal to accurately determine the functional dependencies of the toy system. This becomes obvious when comparing the global model prediction in Figure 3.40 a) to Figure 3.36 b). For this reason, we performed several further iterations of the adaptive design algorithm and activated the confining potential to pull the design points away from the outer edges of the design space.

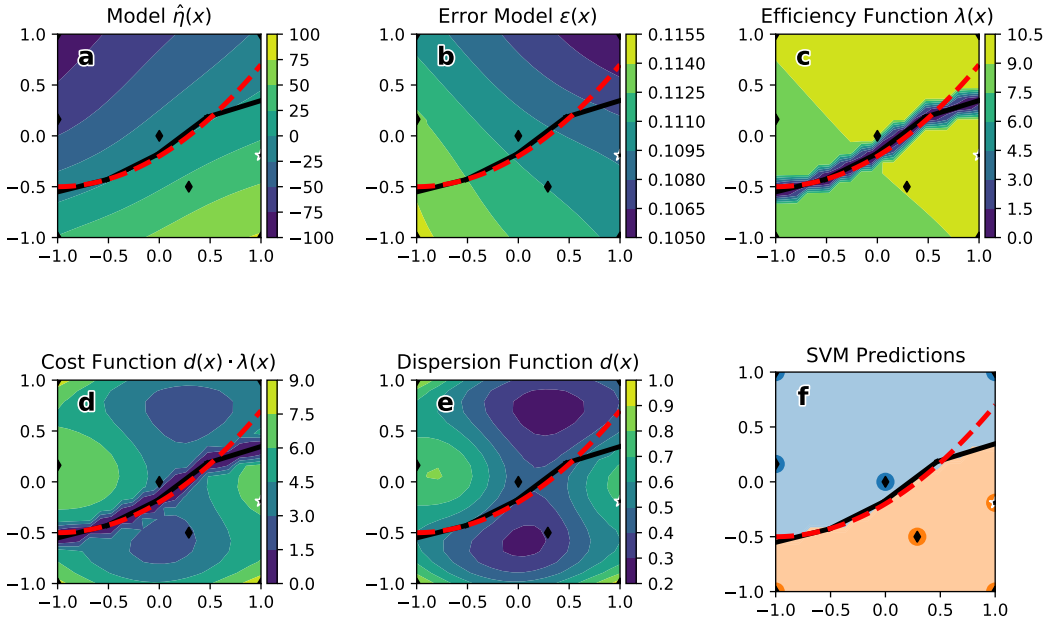


Figure 3.40: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 b). All functions are evaluated after three iterations. Subplots are the same as for Figure 3.30.

The final design after 15 iterations is shown in Figure 3.41. We can see that the position of the regime boundary did not change significantly as the previously determined phase diagram was already close to the known reference. A more pronounced change can be observed for the global response model in Figure 3.41 a), which now shows a more distinct behavior of the two regimes. Further, this extended design provides enough information to fit an accurate error model (Figure 3.41 b)), which captures the dependence of the measurement error on the position of the phase transition.

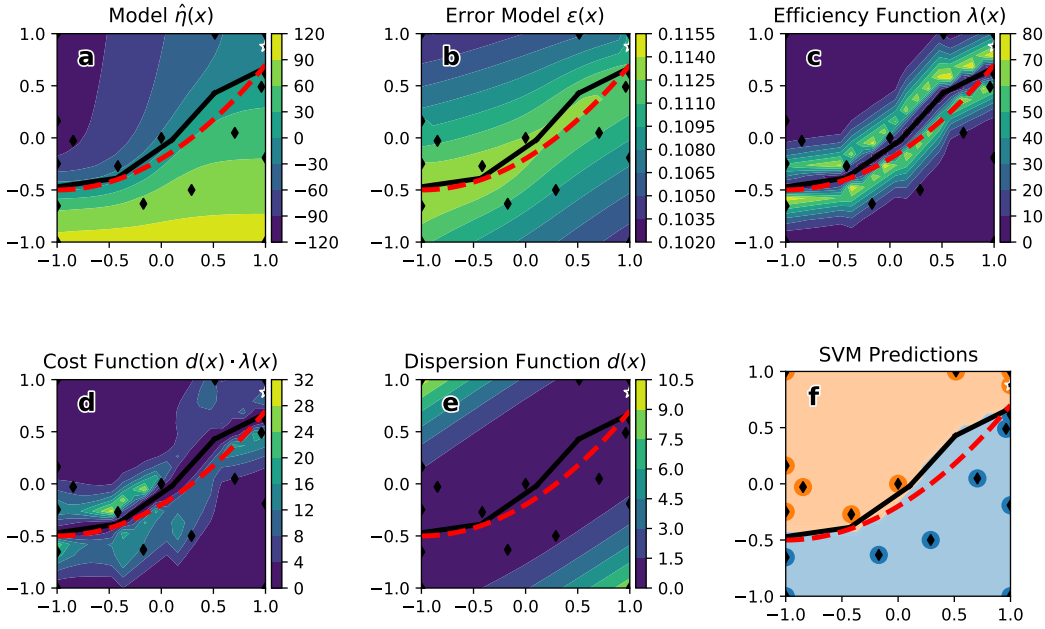


Figure 3.41: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the toy system shown in Figure 3.28 b). All functions are evaluated after 15 iterations. Subplots are the same as for Figure 3.30.

3.3.4 CO Oxidation on RuO_2 Revisited

Based on various simplified toy systems, we could show that the proposed adaptive design algorithm is capable of identifying changes in effective system behavior and the position of the resulting regime boundaries in an efficient way. Coming back to applications in chemical reaction kinetics, we now apply this adaptive design to the CO oxidation on RuO_2, which, in section 3.2, has been investigated based on QMC sampling. For the local model representation of this reaction kinetic system, SPTV models have been used. Therefore, the clustering of the design points is performed on a data set containing both the LASSO coefficients of the local models as well as the corresponding transformation parameters. Again, the final design matrices are given in the appendix A.4.

Kinetic Phase Transition at Constant Temperature

First, we investigated the constant temperature case, resulting in a 2D design space depending on the partial pressures of CO and O_2. We use a 2-Level FFD with additional center point for the initial design. The same experimental plan is used for the local designs around each global center. After the initialization procedure, one new design point is added resulting in two three point regime designs. The corresponding overview is shown in Figure 3.42.

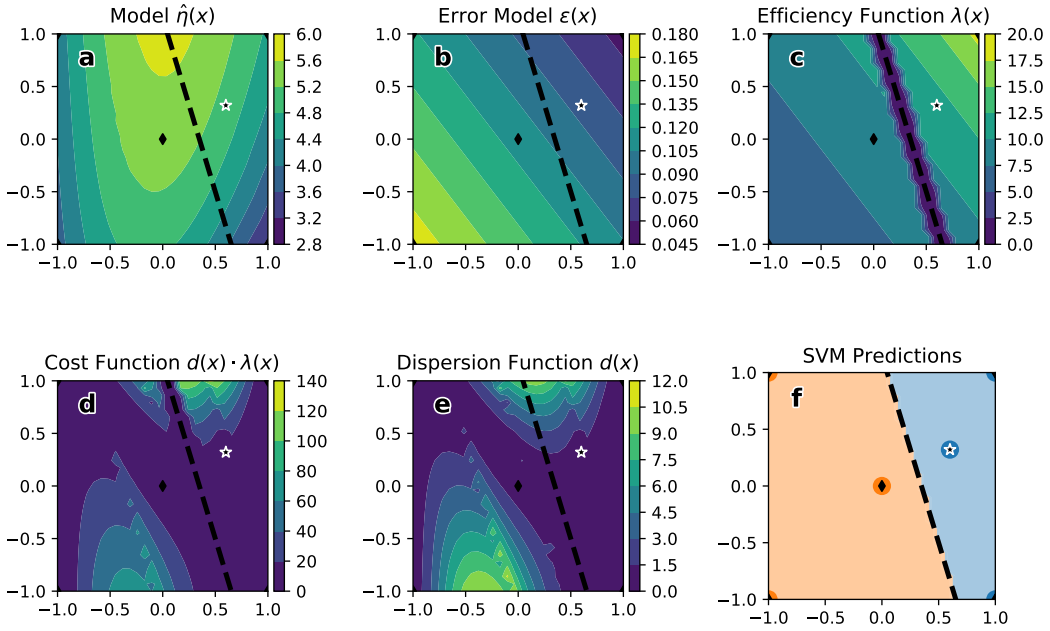


Figure 3.42: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the isothermal CO oxidation over RuO_2. All functions are evaluated after the initialization step. Subplots are the same as for Figure 3.30.

As we have seen for the toy systems, the initialized design allows for setting up a valid surrogate model. The assignment of the design points to different regimes is reasonable. However, the choice of the position of these initial points does not lead to the expected slope of the estimated transition boundary. Additional points are needed in order to correct for this. After two iterations of the Fedorov algorithm, the position of the regime boundary has changed as shown in Figure 3.43. To more accurately determine the position of the boundary, again, the confining potential has been activated and three further iterations have been performed. The resulting design after five iterations is given in Figure 3.44. As expected, the final two points added to the design are located close to the SVC boundary. We can see that the adaptive design successfully identifies the expected kinetic phase transition between the oxygen and the CO covered regimes. Compared to the QMC data set in the previous section, the point density for this adaptive design is not sufficient to identify the transition region as an individual regime. Rather, the identified decision boundary runs through this region of the kinetic phase diagram. Looking at the surrogate model predictions based on this 12 point sequential design in Figure 3.45, we can see despite consisting of only two regimes, the qualitative behavior of the TOF for this reaction is captured (cf. Figure 3.15). The correlation plot shows a systematic overestimation of the reaction rate. This is due to a larger uncertainty in terms of the transition region leading to a broadening of the maximum TOF region as can be seen in Figure 3.45 c). Apart from this overestimation the two regime surrogate model is in reasonable agreement with the MKM taking into account the small number of only 12 training samples.

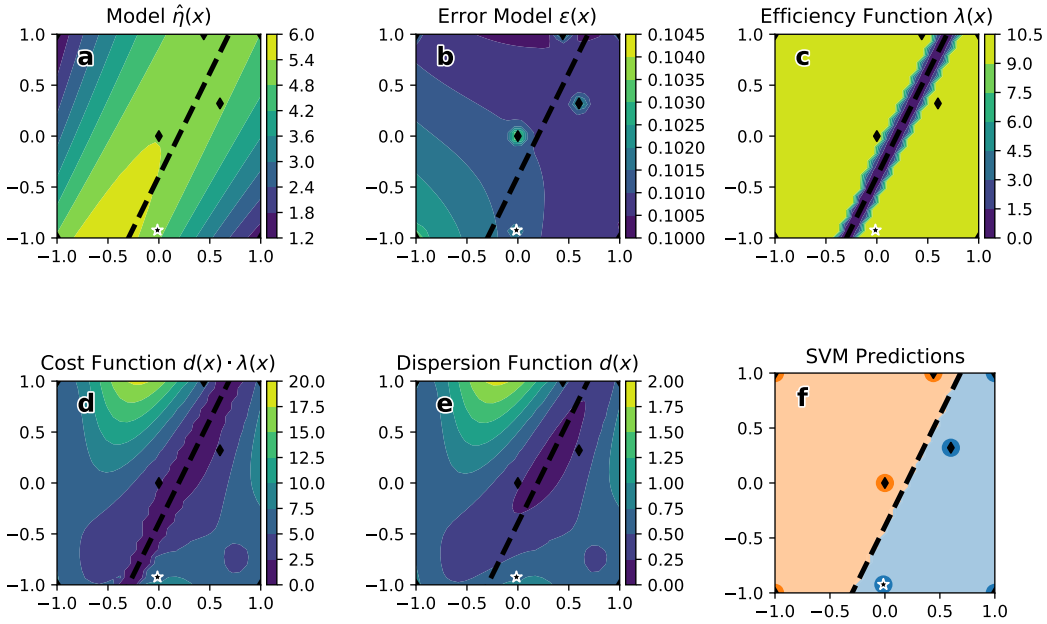


Figure 3.43: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the isothermal CO oxidation over RuO_2. All functions are evaluated after two iterations. Subplots are the same as for Figure 3.30.

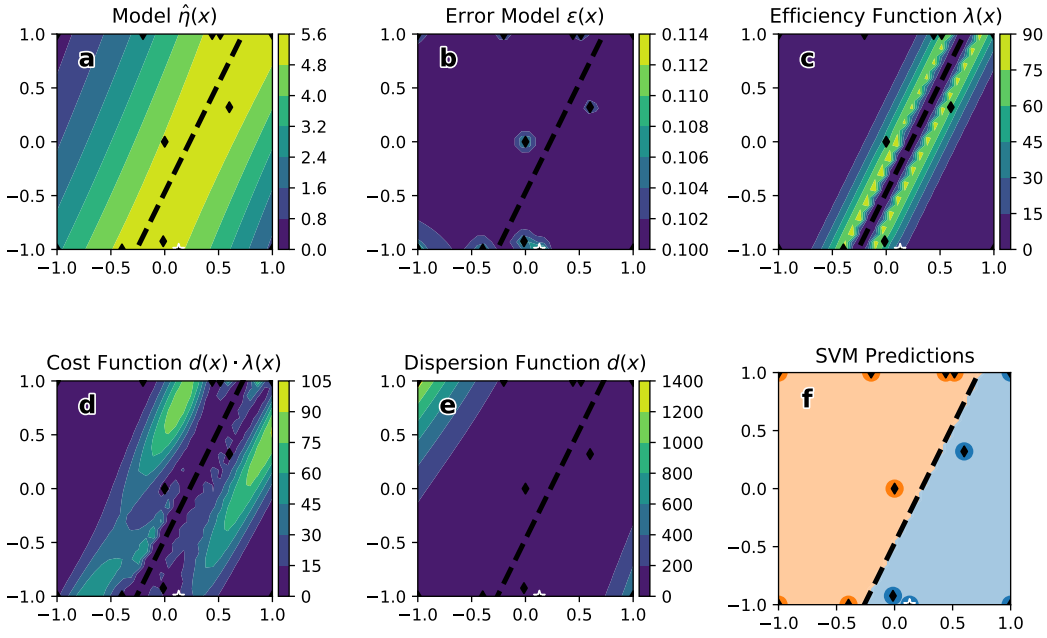


Figure 3.44: Overview over the relevant functions contributing to the optimization problem in the Fedorov algorithm for the isothermal CO oxidation over RuO_2. All functions are evaluated after five iterations. Subplots are the same as for Figure 3.30.

In case the identification of an extended transition region as a separate regime is desired, sampling additional points with a slightly broadened confining potential could be a viable way of determining additional points in this region. However, considering only two regimes

for the surface phase diagram is also a valid description. The transition region resembles reaction conditions where both, the oxygen and CO dominated processes contribute to the overall kinetics. Given a sufficiently dense sampling, points within this region will tend to form their own cluster, as they can not be assigned to either of the major regimes. This, however, is not necessary to capture the effective behavior, as the membership probabilities predicted by our classification model can take care of this mixing of the two kinetic processes over the phase transition.

We compared the position of the decision boundary based on the 12 point sequential adaptive design (SAD) to several sets of QMC points. Also included are the result for a sequential design run using the batch optimization procedure with a batch size of three points per iteration. For a simpler comparison we restricted ourselves to a single linear regime boundary, which is uniquely defined by a slope and an intercept. The results are given in Table 3.8.

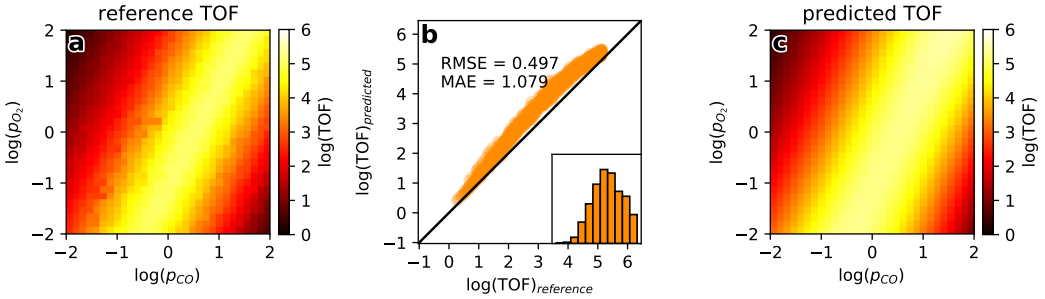


Figure 3.45: Prediction performance of the global surrogate model based on the final 12 point adaptive design evaluated on a regular grid over the entire reaction condition domain. a) TOF calculated using the reference MKM. b) Parity plot between reference and predicted TOF values. The parity plot as well as the error values are determined on the test set depicted in the other subfigures. This test set is constructed as a regular 30 by 30 grid. The inset shows the shape of the residual distribution. Despite not being centered around 0, the distribution is reasonably symmetric. c) TOF predicted by the global surrogate model.

Table 3.8: Parameters of the regime boundary for the two regime problem of the CO oxidation on RuO_2 for different designs.

Design	# Points	Intercept	Slope
SAD	12	-0.47	2.03
SAD_{batch}	13	-0.86	2.72
QMC(Halton)	12	-0.69	2.67
$QMC(Halton)_{flipped}$	12	-1.00	3.38
QMC(Sobol)	12	-0.75	3.00
$QMC(Sobol)_{flipped}$	12	-0.75	3.00
Reference(LHC)	120	-0.67	2.18

As the position of the real phase transition is difficult to define, we used an extensive latin hypercube (LHC) sampling with 120 points as reference. Comparing slope and intercept values of the decision boundary obtained by our adaptive design to QMC designs of similar size, for most cases, the adaptive design is significantly closer to the extensive reference data set. One exception is the performance of the 12 point Halton sequence, which yields comparable results to our approach. As QMC sequences are deterministic, these 12 points will always be the same, independent of the position of the phase transition. If the

transition happens to fall in between the points of such a QMC sequence, the result may compare to or even outperform the adaptive design. However, it is not guaranteed that the phase transition is well sampled for any system. Table 3.8 also shows the result for a flipped Halton sequence, were we change the assignment order of the two partial pressures. We can see that the resulting decision boundary deviates significantly more from the LHC reference. This is an indication that the good performance of the original series was indeed by chance. As opposed to the Halton series, Sobol QMC samples are symmetric with respect to an interchanging of the two axes. This property of the Sobol sequence leads to both the original and the flipped sample resulting in the same position of the decision boundary. Comparing the single point SAD to the batch optimization, we can see that the batch SAD deviates slightly stronger from the reference, especially in terms of the intercept. This behavior is expected, as the fully sequential optimization takes into account the full information of the current estimate in every iteration. Instead, for the batch case, multiple minima of a fixed cost function are sampled, which could have changed, if the points would have been measured sequentially. Still, the optimization identifies the right trend. The batch optimization also has the advantage, that it requires only two iterations to construct this 13 point design, which could be beneficial in a real experimental study. The batch optimized design contains 13 points as opposed to the sequential one with 12, as the initialization procedure for the batch SAD added one additional point. Continuing with a fixed batch size results in the slightly larger final design.

Hierarchical Approach for the Exploration of Higher Dimensional Problems

The regime identification formalism using a modified Fedorov algorithm, we propose in this work, can easily be extended to higher dimensional problems. Given an N dimensional design space, the corresponding regime boundaries are (N-1)D subspaces. A third experimental factor in the CO oxidation on RuO_2 is temperature. By varying temperature in addition to the partial pressure of both reactants, we span a 3D design space. Hence, the resulting regime boundaries are 2D planes or, generally, 2D surfaces. We applied the adaptive design algorithm to the temperature dependent problem by initializing the algorithm with a three factor Placket-Burman design (PBD) plus an additional central point:

$$D_{initial} = \begin{bmatrix} -1 & -1 & +1 \\ +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & +1 & -1 \\ +1 & -1 & -1 \end{bmatrix}. \tag{3.23}$$

Starting from this five point designs, the initialization procedure has been performed to ensure locally nonsingular designs. The evolution of the adaptive design and the resulting regime boundary over several iterations is illustrated in Figure 3.46. The minimum sample requirement for a 3D linear model is four. As we can see, the design initialization step added six points to the initial PBD in order to fit the global surrogate model. This indicates major changes in the position of the SVM decision boundary during this phase. The following sample points proposed by the adaptive design mainly focus on the corners of the design space reproducing classical factorial experimental designs. As soon as most corners of the design space are covered, the algorithm adds points along the edges of the 3D cube to more accurately localize the phase transition. This goes along with few points within the design space volume, which facilitate a more accurate estimation of the regime model parameters. We can see that already from the second iteration onward, the position of the regime boundary does no longer change significantly.

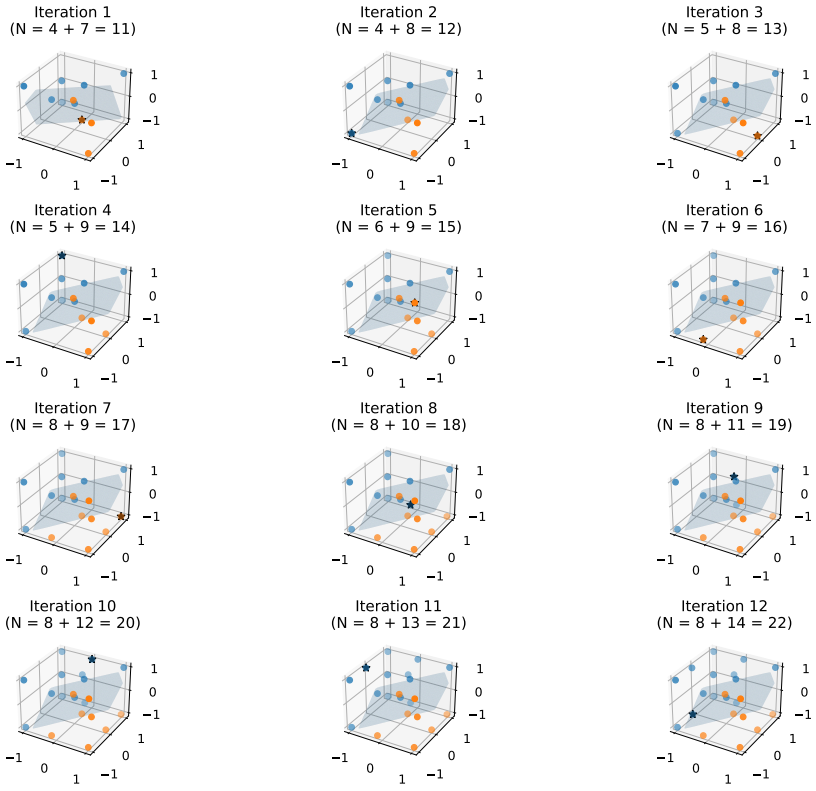


Figure 3.46: Adaptive design algorithm for the 3D temperature dependent CO oxidation on RuO_2. The subplots show the design points as well as the 2D decision boundary for every iteration. The star indicates the latest addition to the design.

Having identified a 2D subspace of the 3D design space which resembles the transition region and hence the region of maximum TOF, we can map a 2D experimental design onto this plane to efficiently fit a more accurate response model for this region. From the parameters of the decision boundary determined by the SVC we can directly obtain a coordinate system lying within this plane. However, in mapping onto this plane the question of the design space boundaries arises. While the original 3D design space was a well defined (hyper-)cube, the intersection of this cube with the SVC decision boundary is in general a simplex with ≥ 4 corners. In order to work on a simple rectangular 2D subspace, we, therefore, need to determine the proper rotation of the in plane coordinate system, such that the inscribed rectangle of the simplex has a maximum area. Given a rotation angle around the normal vector of the plane, we can identify the maximum inscribed rectangle using an algorithm described by Marzeh et al.[136]. For the 2D case, this algorithm is efficient enough, such that the possible rotations can be evaluated brute force on a regular grid. The resulting mapping from the original 3D space onto the rotated 2D subspace is illustrated in Figure 3.47.

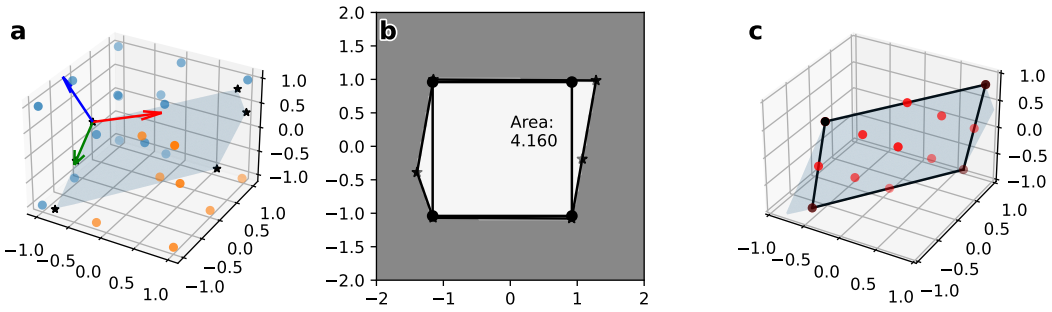


Figure 3.47: Definition of the 2D coordinate system located in the decision boundary plane. a) The 3D design points and the corresponding decision boundary. The stars indicate the intersections of the plane edges with the overall boundaries of the design space. The three vectors indicate the basis vectors of the coordinate system within the plane with one dimension being normal to the plane. b) Maximum area rectangle enclosed by the simplex defined by all intersection points of the decision plane with the global boundaries. The x- and y-axis correspond to vectors within the plane. c) A combination of a 3-Level FFD and 2-Level FFD mapped onto the decision plane used for training a response model in the resulting 2D coordinate system.

Using the maximum area rectangle as the new 2D design space, we can construct a 2D experimental design within this plane in order to fit a response model of the maximum TOF region. The corresponding sample points are shown in Figure 3.47 c). The response model was fitted as a function of the original design parameters, the partial pressures of CO and oxygen as well as temperature. Figure 3.48 shows the prediction quality of the resulting model.

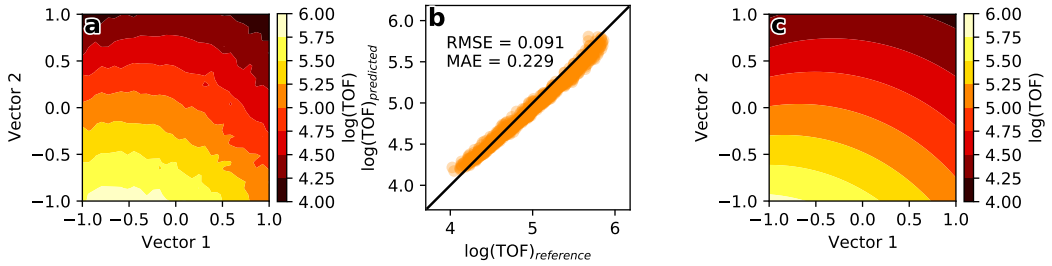


Figure 3.48: Prediction performance of the decision plane surrogate model fit to the 2D experimental design in Figure 3.47 c) evaluated on a regular grid over the entire reaction condition domain. a) TOF calculated using the reference MKM. b) Parity plot between reference and predicted TOF values. The parity plot as well as the error values are determined on the test set depicted in the other subfigures. This test set is constructed as a regular 30 by 30 grid. c) TOF predicted by the global surrogate model.

The mapping between the 2D subspace within the boundary region and the original coordinates is a simple linear transformation. The coordinate system is rotated such that two of the three basis vectors are located within the transition plane. This is followed by a rotation around the normal vector of the plane to maximize the area of the resulting rectangle. By storing the original intercept of the rotated 3D coordinate system before the projection to 2D, the entire mapping is easily invertible.

The basis vectors of the transition plane in terms of the original coordinates are the following:

$$\begin{aligned} \text{Vector 1} &= \begin{pmatrix} -0.200 \\ 0.145 \\ -0.969 \end{pmatrix} \begin{matrix} \log(p_{CO}) \\ \log(p_{O_2}) \\ \hat{T} \end{matrix} \\ \text{Vector 2} &= \begin{pmatrix} -0.569 \\ -0.822 \\ -0.005 \end{pmatrix}^T \begin{matrix} \log(p_{CO}) \\ \log(p_{O_2}) \\ \hat{T} \end{matrix} \end{aligned} \tag{3.24}$$

We can see that one of the two basis vectors (Vector 1) is heavily dominated by temperature and shows only a slight tilt in the other two dimensions. Vector 2 on the other hand is largely independent of the reaction temperature and depends on both partial pressures. This behavior has already been observed for the extensive QMC sampling in the previous section. The position of the transition region between the oxygen and the CO poisoned regimes does not show a large shift due to temperature effects. In fact, the dependence of the TOF on temperature does not show an abrupt phase transition within the range of reaction conditions investigated here. Using the proposed adaptive design algorithm, this simplified coordinate system, which can be used in more detailed kinetic studies, could be identified without any prior knowledge of the underlying microkinetics.

Finally, we again compared the position of the estimated phase transition of the SAD with QMC point sets of the same size. As we are now dealing with a 2D transition plane, we need to assess the similarity of two regime boundaries using a different metric. The two quantities of interest here, are the deviation in euclidian distance of the plane from the origin and the scalar product of the two normal vectors. If the normal vectors are scaled to unit length, the absolute value of this scalar product is connected to direction cosine of the two vectors and takes on values between 0 and 1. A value of 1 would correspond to two parallel planes, while a scalar product of 0 would indicate that the normal vectors of the two planes are orthogonal. A 200 point LHC design acts as the reference for this comparison and the results are summarized in Table 3.9.

Table 3.9: Parameters of the 2D regime boundary for the temperature dependent 3D two regime problem of the CO oxidation on RuO_2 for different designs.

Design	# Points	Normal Vector	Scalar Product	Δ Origin Distance
SAD	22	$\begin{pmatrix} -0.798 & 0.550 & 0.247 \end{pmatrix}^T$	0.993	0.971
QMC(Halton)	22	$\begin{pmatrix} -0.955 & 0.251 & 0.158 \end{pmatrix}^T$	0.969	1.763
QMC(Halton)$_{\text{flipped}}$	22	$\begin{pmatrix} -0.886 & 0.352 & 0.300 \end{pmatrix}^T$	0.995	1.380
QMC(Sobol)	22	$\begin{pmatrix} 0.784 & -0.510 & -0.353 \end{pmatrix}^T$	-0.992	3.371
Reference(LHC)	200	$\begin{pmatrix} -0.851 & 0.446 & 0.277 \end{pmatrix}^T$	1.000	0.000

Comparing the SAD to the LHC reference the normal vectors of the two resulting transition planes are almost perfectly parallel with a scalar product of 0.993. Of all investigated designs consisting of 22 points, the SAD also shows the lowest distance in terms of the plane's intercept. Looking at the performance of the QMC samples, we can see that both Sobol and Halton sequences result in very accurate normal vectors with the Halton sequence even outperforming the SAD. However, in terms of the intercept values, the distance is, again, larger as compared to the sequential design. For the Halton sequence, we again switched the assignment of the experimental factors to the sample dimensions. The resulting decision boundary is in closer agreement with the reference compared to the original Halton points, however, as for the 2D case this improvement is purely by chance, as the QMC design does not adapt to the phase transition.

3.3.5 Outlook: Applications in Real Experiments

We have shown a sequential adaptive design algorithm specialized for the identification of regimes in reaction kinetic systems. Building on the original adaptive D-optimal design procedure proposed by Fedorov [11], we have introduced the presence of regime boundaries, which mark an abrupt change in effective behavior of the system. This is incorporated into the global model assumption by constructing the information matrix based on a coupled SPTV model for the multi-regime system. At the same time, we modify the measurement efficiency function to avoid putting samples exactly on the regime boundary, where the expected noise level is high and a safe assignment of the sample to any regime can be difficult. Looking at various artificial toy systems as well as the MKM of the CO oxidation over RuO_2, we could show, that this algorithm accurately determines the location of such phase transitions in an efficient way.

This data efficiency becomes especially important when applying such a procedure in a real experimental study, where every single design point is extremely expensive to evaluate. A major drawback of this procedure for experimental studies is the way the local fingerprint for the effective kinetics is obtained. For every global design point, in our algorithm, a local factorial design is expanded in order to estimate the local sensitivity of the kinetics. These local designs lead to a dramatic increase in experimental effort, if every individual factor setting is equilibrated in a reactor. A possible solution to this problem could be frequency response measurements [137]. Such experiments allow for an estimation of kinetic parameters based on small periodic modulations of experimental conditions around the steady-state [138, 139]. Replacing the local fingerprint used in the regime assignment procedure by such frequency response estimated parameters would effectively reduce the number of required reactor equilibrations to the number of global design points. For example, in the case of the isothermal CO oxidation over RuO_2, reaching a steady-state for 14 different reaction conditions determined by the SAD would be sufficient.

In general, the here proposed adaptive design method is not restricted to the investigation of reaction kinetic phase diagrams. Only minor modification would be required to generalize this approach for arbitrary systems expected to show phase transitions of some kind. Using SPTV models as a representation of the regime wide models is of course tailored towards reaction kinetics problems. However, in principle every kind of regression model could be used instead. In the most general case, as shown for the artificial toy systems, standard polynomial regression models can be employed, which tend to have reasonable approximation properties in sufficiently smooth regions. Going in the direction of catalyst design, this method could also be applied to identify phase transitions in a design space spanned by e.g. the synthesis conditions of the catalyst. This way, expensive screening and optimization studies could be restricted to a smaller parameter range or lower dimensional manifolds allowing for a more efficient sampling.

Up to this point the application test cases for the algorithm in this work have been rather simple in terms of the kinetic phase diagram and the smoothness of the in-regime behavior. It will be interesting to benchmark the performance of this algorithm for more noisy data. Generating kinetic data based on kMC [4] could turn out as a challenge for the modeling algorithm. Also, the presence of many regimes in a system needs to be further investigated, as the various penalty terms may lead to an insufficient sampling of especially smaller regimes. Further, while the formulation of the method allows for a straight forward extension to higher dimensional problems, some open questions still need to be addressed. For example, the piecewise linear approximation of nonlinear regime boundaries is currently only implemented for 2D two regime systems. One of the problems here is to identify the proper decision function thresholds to locate points close to the decision boundary in a multi-class problem. To this end, selecting points based on the local uncertainty in terms of the calibrated membership probabilities of the SVC could be a valid alternative. Further, when selecting points based on such a threshold, samples along all decision boundaries are selected. To approximate multiple regime boundaries individually, these points would need to be separated, a task which could for example be tackled by clustering analysis.

Chapter 4

Summary, Conclusions and Outlook

"Das ist wie Schiffe Versenken, nur hochdimensional." - Christoph S.

Understanding the kinetic phase diagram of chemical reactions is central in catalysis research. The complex mechanisms of surface reactions give rise to distinct kinetic regimes, which show qualitatively different behavior. These regimes are separated by sharp kinetic phase transitions [34], which generally coincide with transitions in the surface phase diagram of the catalyst [3]. Knowledge about the kinetic phase diagram of a catalytic reaction is valuable when trying to identify optimal conditions in a reactor setup. In this thesis, we have presented a data driven top down approach to identify the shape of kinetic phase diagrams based on experimental observations.

In a first step (section 3.1), we developed an algorithm to identify effective kinetic models from kinetic data. This algorithm approximates the kinetics of the system as a sparse polynomial in terms of the reaction conditions. These factors entering the polynomial expansion are taken to their respective sensitivity parameter with respect to the response variable. In general this response variable is identified with the observed rate of the reaction and thus for example the sensitivity parameter for a given reactant concentration can be interpreted as the corresponding partial reaction order. The resulting models represent the best analytical mean field approximation to the observed kinetics and take the shape of effective rate laws known from classical kinetics. In order to estimate the model parameters for these rate laws, we employed modern statistical and numerical methods. Constructing the polynomial models making use of the LASSO [6, 45] ensures sparsity of the final rate law, as only those terms are selected from the polynomial ansatz, which contribute in a significant way. Further, also the nonlinear sensitivity parameters are optimized for these models. In a multi-objective framework [98] we employ a genetic algorithm [8] to identify the best possible trade off between the accuracy of the kinetic model and the adequacy of the normality assumption on the residual distribution [7]. A key requirement for the estimation of interaction terms in the polynomial ansatz is the proper factorial design of the kinetic data sets. We could show that this sparse approximation approach combined with classical DoE reproduces rate laws of classical kinetic reaction networks [80]. Investigating a MKM for the CO oxidation on a Pt catalyst, we could further exemplify the capability of this approach to distinguish between effective kinetic regimes.

In order to locate the position of a kinetic phase transition as a function of various process parameters, we harness the capability of this novel kinetic modeling approach to automatically identify kinetic rate laws (c.f. section 3.2). In an automatized regime identification procedure, we assign sample points to different regimes by analyzing the behavior of local kinetic models. We sample the range of possible reaction conditions making use of QMC sequences [68] and expand local factorial designs around each of these QMC centers. Based on the local designs, kinetic models are optimized and the behavior of these models is encoded in a vector of model parameters. On the data set of these kinetic fingerprints, we peform a clustering analysis [99] in order to label each QMC center. We could show that this clustering according to the kinetic fingerprints of the samples also results in a reasonable

segmentation in the reaction condition space. Based on the labels of each QMC center, a classification model [9] is trained. The decision boundaries of this classification model then correspond to an approximation to the real kinetic phase transitions. For the CO oxidation on RuO_2, the support vector classifier trained following this procedure nicely reproduced the catayst's surface phase diagram from the reference MKM. Going one step further, we could establish that membership probabilities from a calibrated classification model [76] can act as weights in a linear combination, coupling the local regime models to form a global representation of the multi-regime kinetics.

Working towards applying this regime identification procedure in real experimental studies, we investigated adaptive design methodologies (c.f. section 3.3). Adaptive designs are experimental plans which are updated in every iteration making use of the information obtained during the previous measurements. Such computer optimized designs require some assumption on the functional form of the model we want to use in order to describe the system at hand. Using coupled local effective models, we incorporate the presence of regime boundaries into this model assumption and thus into the optimality criteria for the experimental design. While the QMC sequences sample the design space as evenly as possible, the adaptive design takes into account the current guess for the position of the phase transition when suggesting another design point. This way, the exact shape of the kinetic phase diagram can be explored in an iterative manner. For several analytical toy systems, using this algorithm, the position of regime boundaries could accurately be reproduced. Further, revisiting the CO oxidation on RuO_2, we could show that this algorithm can also capture kinetic phase diagrams of moderately complex reaction networks. For higher dimensional problems, we demonstrated how to reduce the number of effective coordinates by projecting lower dimensional designs onto the boundary regions identified by our algorithm.

Having exemplified the performance of our algorithm based on artificial data sets, we envision the application of this kind of technique in the investigation of novel and mechanistically not well understood reaction systems. In the context of computer guided experimentation, such an adaptive design algorithm can pave the way for an almost automatic exploration of kinetic phase diagrams. Provided a set of process parameters and an initial design, the algorithm can guide the experimental search through the kinetic space. In the process of this exploration, efficient surrogate models are obtained, which can be utilized for optimization or coupled to macroscopic simulations in a reactor engineering framework.

From an algorithmic point of view, the presented adaptive procedure could also pave the way towards handling continuous and categorical factors on an equal footing in a design framework. The piecewise constant toy models covered in section 3.3.3, can be interpreted as a continuous representation of a design space depending on categorical variables, where each regime corresponds to a constant factor setting. Often times, declaring categorical variables is a low order approximation to an actually more complex problem. In such cases, taking into account the less idealized small variations around points within the same regime (categorical setting), which show qualitatively similar behavior, may actually be beneficial for locating the transition point of the pseudo-categorical factors.

Acknowledgements

Zuerst möchte ich mich bei Professor Karsten Reuter dafür bedanken, dass er mir die Durchführung meiner Doktorarbeit in seiner Gruppe ermöglicht hat. Mit wertvollem Input trug er maßgebend zum Erfolg dieser Arbeit bei, besonders wenn es darum ging die Ausrichtung des Projekts zu überdenken und Resultate zu kommunizieren. Darüber hinaus, konnte er, nicht zuletzt dank der regelmäßigen Workshops und Events sowohl an der TUM, als auch am FHI ein unglaublich angenehmes Arbeitsumfeld schaffen.

Ein ebenso großer Dank geht an meinen Betreuer, Christoph Scheurer. Ohne Christoph wäre diese Arbeit nicht möglich gewesen. Mit seiner ausdauernden Ruhe, konnte er zahlreiche Male über mangelnden Durchblick meinerseits hinwegsehen und mir in vielen Diskussionen Fragen zu beliebigen Themen beantworten. Sollte er mal etwas nicht parat haben, wusste er aber auch direkt aus dem Stehgreif die entsprechende Literaturstelle ("Das müsstest du im Box & Draper in Kapitel 8 nachlesen können..."). Seine brillanten Ideen und seine mathematische Herangehensweise haben diese Arbeit geprägt und mir eine ganz neue Denkweise nahegelegt.

Die letzten Jahre wären nicht die Gleichen gewesen, ohne die tollen Kolleginnen und Kollegen in dieser Gruppe. Ich möchte mich bedanken bei Hanna, Carsten und Simon, der Crew aus dem Mystery Room, die lange den alten Trakt in Garching unsicher gemacht hat, bei meinem PostDoc Julius, bei der Gruppe um Thorben, Sina, Simon, Elisabetta und Nam, die den ersten Brückenkopf in Berlin aufgebaut haben, sowie bei der IT-Crew um Matthias, David, Christoph, Christian und Simeon. Auch bedanke ich mich bei Ruth, Julia und Steffen, die den Laden im Hintergrund am Laufen halten. Genauso möchte ich mich bei allen, deren Namen hier nicht explizit genannt wurden für die hilfsbereite und freundliche Atmosphäre bedanken, für die man immer wieder gerne ins Büro gekommen ist.

Ich möchte mich bei MuniCat und Clariant für die finanzielle Unterstützung meiner Arbeit im Rahmen des MuniCat Projekts ISS4HC bedanken. Mein Dank gilt auch den Kollegen von Seiten der TUM und Clariant für die Zusammenarbeit und den regelmäßigen Austausch.

Ein riesiger Dank gilt meinen Eltern Angelika und Wolfgang und meinen Geschwistern, Celine, Moritz und Joscha, die mir meine Ausbildung ermöglicht haben und mir dabei mit Rat und Tat zur Seite standen. Ich bedanke mich bei meinen Freunden, die mir durch Freizeit, Sport und Rumblödeln einen wichtigen Ausgleich zur Arbeit geboten haben. Zu guter Letzt, möchte ich mich bei Franziska bedanken, die vehement versucht hat, mir das Konzept Feierabend nahe zu bringen, sich aber auch beim Frühstück wirre Dinge über Pareto-Fronten angehört hat. Danke für deine Unterstützung und Geduld!

Appendix A

Appendix

A.1 Additional Details on the Regression Algorithm

The contents of this section have been published in the Chemical Engineering Journal under the CC-BY license [1] (cf. SI) .

A.1.1 Influence of the Noise Term

The algorithm is designed to identify robust models based on "real" data from experimental measurements. Such data is subject to statistical noise, which in the limit of a large number of measurements can be assumed to be normally distributed. For this reason, all simulated rates based on simple mikrokinetic models have been modified by adding a noise term from a normal distribution. This, to a certain degree, is necessary, as the algorithm tries to find the model with most gaussian shaped residual distribution. If we leave out this term, the main source of error would come from the discretization when estimating the rate from finite differences. This discretization error, however, is in general not normally distributed [140]. Figure A.1 shows how the magnitude of the noise influences the result of the pareto optimization and the corresponding noise standard deviations are given in Table A.1. We can see, that up to a noise level of 10% of the mean data value, the trends in transformation parameters are robust. Also, we see that even without artificial gaussian noise, the correct reaction orders are captured. The error in our simulated data set thus seems to be already close to a normal distribution. The additional noise term we add is therefore a safety measure to assure, the underlying assumptions for our algorithm are given. The assumption, that experimental noise is approximately normally distributed in the limit of large data sets on the other hand is well established due to the central limit theorem [102].

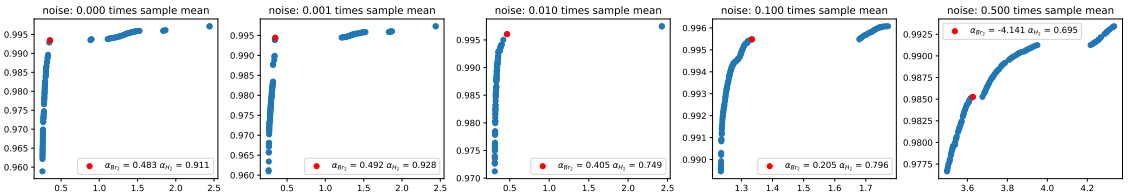


Figure A.1: Pareto fronts for the isothermal HBr problem at different noise levels. The red points indicates a reasonable choice for the final model with the transformation parameters given in the legend. The random errors are drawn from distributions with mean 0 and standard deviations given in Table A.1.

Table A.1: Standard deviations for the different noise levels.

Percentage of sample mean	Standard deviation [mol/l/s]
0.0%	0.000
0.1%	$2.516 \cdot 10^{-8}$
1.0%	$2.516 \cdot 10^{-7}$
10.0%	$2.516 \cdot 10^{-6}$
50.0%	$1.258 \cdot 10^{-5}$

A.1.2 Model Extrapolation

Figure A.2 shows the extrapolation of maximum region models produced with different zero shift values.

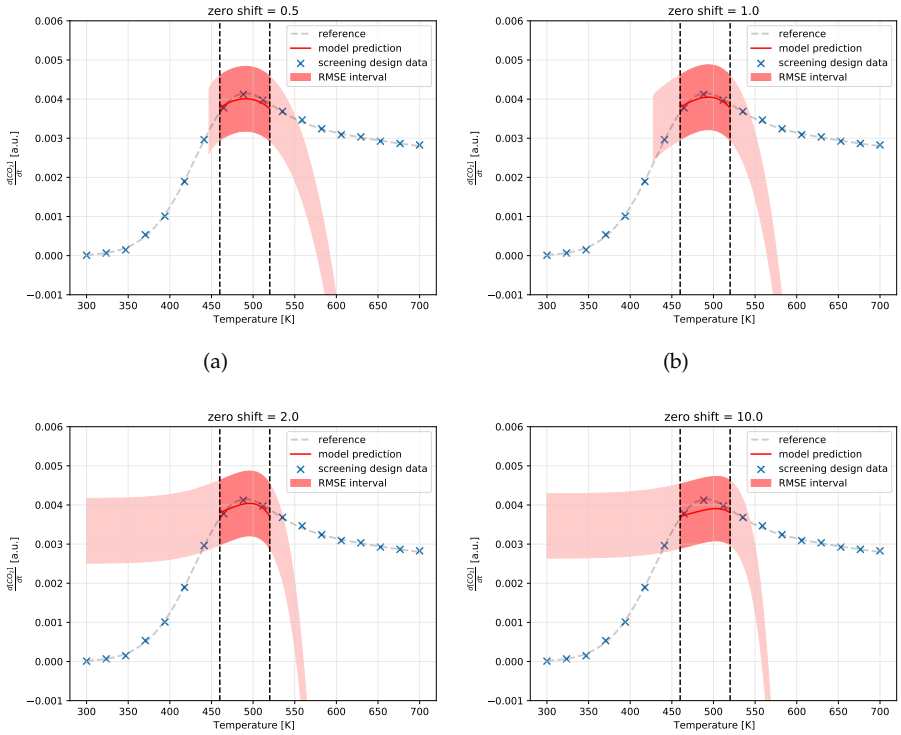


Figure A.2: CO oxidation models with different zero shift values.

It immediately becomes obvious, that the extrapolation beyond the training region (dashed black lines) is not recommended with such polynomial models. The curvature of the model is solely determined by the training data and might lead to dramatic deviations and does not need to resemble the behavior of the "real" function outside this limited domain. The higher the order of the polynomial, the more drastic these deviations can become. Additionally, the extrapolation to lower temperatures is limited by the the zero shift value used in the preprocessing steps of the algorithm. If the input features end up with negative values after the shifting and scaling steps, taking a logarithm during the power transformation becomes a problem. Thus, for such values no prediction is possible. Increasing the zero shift value resolves this issue, however, at the cost of a decreased sensitivity to the exponents of the transformation, as the absolute values of the features will increase. This will result in a worse fit as can be seen in Figure A.2.

A.2 Additional Results

A.2.1 Temperature Dependent HBr Reaction

As stated in the main text, the chosen temperature range between 840 K and 860 K is extremely narrow. From an experimental point of view, this temperature range would be difficult to resolve due to the limitations of experimental setups. As already mentioned above, the scaling of the temperature input introduces certain limitations regarding the extend of the temperature range of interest. Here we show two additional examples for the temperature dependent HBr reaction. This time, we increased the magnitude of the temperature variations to 100 K. Figures A.3 and A.4 show the results for temperature windows of 900 K to 1000 K and 1300 K to 1400 K respectively. As can be seen from these results, the kinetics of the HBr reaction is qualitatively reproduced also in this larger temperature regimes. The reaction orders and activation barriers also reproduce the expected behavior. Looking at the error distributions of the resulting models, however, they do not resemble a gaussian shape. We attribute this result to the way larger temperature effect on the simulated rates, compared to the effects of the concentration variations. This results in the temperature dominating the LASSO regression. Looking at the parity plot e.g. in Fig. A.3, we can clearly see that the data points are separated into three domains, which correspond to the three sampled temperature values.

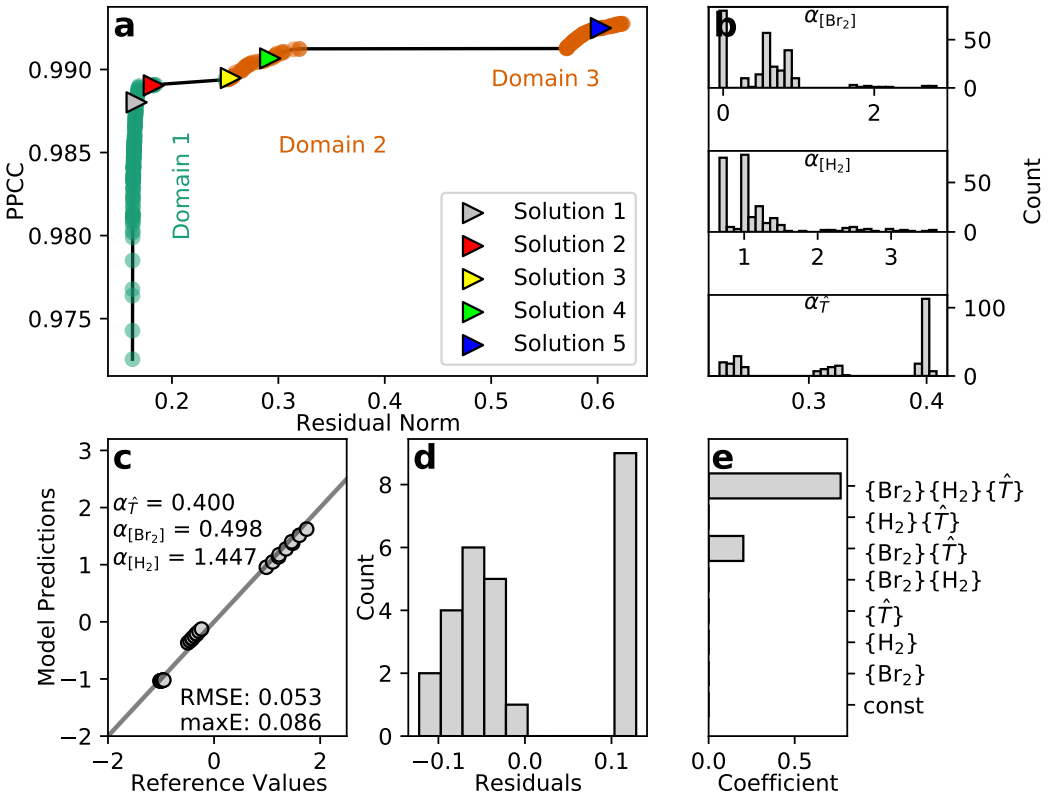


Figure A.3: a) Pareto front for the temperature-dependent HBr problem in the range 900 K to 1000 K. The NSGA-II optimization is performed on a population of 1000 individuals. Five representative pareto-optimal solutions are highlighted with their respective sensitivity parameters given in Table A.2. b) Distribution of reaction orders α for H_2, Br_2 and the scaled temperature values $\hat{T}$ over the entire population. c) Parity plot for solution 1. d) Residual distribution for solution 1. e) LASSO regression coefficients for solution 1.

Table A.2: Model parameters and objective values for five selected temperature-dependent pareto optimal HBr models in the range 900 K to 1000 K corresponding to the pareto front in Fig. A.3.

	$\alpha_{[Br_2]}$	$\alpha_{[H_2]}$	$\alpha_{\hat{T}}$	Residual	PPCC
▶ Solution 1	0.498	1.447	0.400	0.167	0.988
▶ Solution 2	2.701	3.618	0.408	0.183	0.989
▶ Solution 3	0.554	2.133	0.327	0.255	0.989
▶ Solution 4	0.558	3.146	0.314	0.292	0.991
▶ Solution 5	0.006	0.695	0.232	0.603	0.992

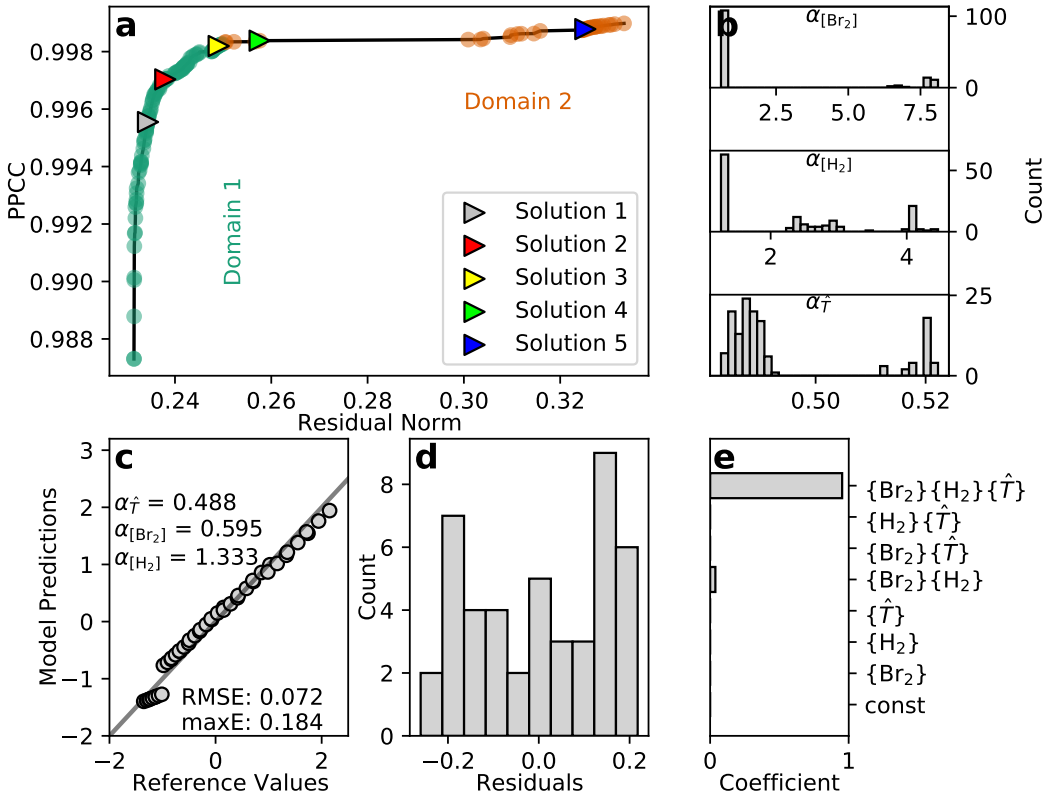


Figure A.4: a) Pareto front for the temperature-dependent HBr problem in the range 1300 K to 1400 K. The NSGA-II optimization is performed on a population of 1000 individuals. Five representative pareto-optimal solutions are highlighted with their respective sensitivity parameters given in Table A.3. b) Distribution of reaction orders α for H_2, Br_2 and the scaled temperature values $\hat{T}$ over the entire population. c) Parity plot for solution 2. d) Residual distribution for solution 2. e) LASSO regression coefficients for solution 2.

A.2.2 Catalytic Cycle without Side Reaction

To observe the change in effective kinetics with an increasing excess in one reactant, we performed multiple simulation runs of the catalytic cycle at different A:B ratios. For further analysis, only one pareto optimal solution was chosen. The complete pareto fronts for the 7 different feed ratios are given in Fig. A.5.

Table A.3: Model parameters and objective values for five selected temperature-dependent pareto optimal HBr models in the range 1300 K to 1400 K corresponding to the pareto front in Fig. A.4.

	$\alpha_{[Br_2]}$	$\alpha_{[H_2]}$	$\alpha_{\hat{T}}$	Residual	PPCC
▶ Solution 1	0.611	1.307	0.486	0.234	0.996
▶ Solution 2	0.595	1.333	0.488	0.238	0.997
▶ Solution 3	0.607	2.919	0.491	0.249	0.998
▶ Solution 4	0.645	3.424	0.492	0.258	0.998
▶ Solution 5	7.699	4.091	0.520	0.325	0.999

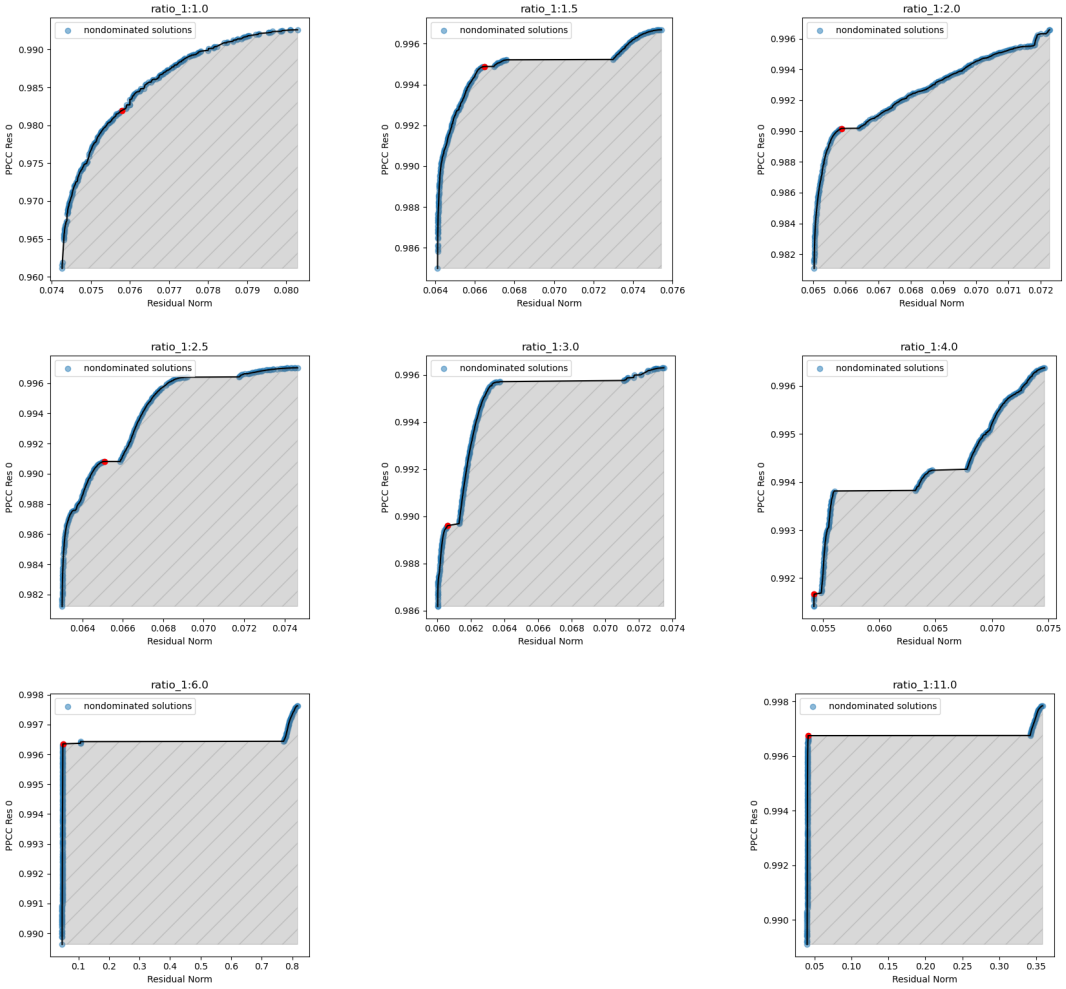


Figure A.5: Pareto fronts for the catalytic cycle without side reaction at different A:B ratios. The red point indicates the selected solution.

The underlying data sets have been obtained by sampling minor variations of the two reactant concentrations according to 2 factor full factorial design (25 runs) with level settings given in Table A.4. All runs have been performed at a constant temperature of 700 K.

A.2.3 Catalytic Cycle: Substrate Decomposition

Figure A.6 shows the pareto front for the isothermal catalytic cycle with substrate decomposition. As two response variables (rate towards P and rate towards D) are considered, the cost function is no 3 dimensional. The automatic pareto front model selection, can in

Table A.4: Parameter scalings for the catalytic cycle without side reaction. x indicates the excess in reactant B and takes values of 0, 0.5, 1, 2, 4, 6, 10.

Level	[A] [1·10^{-4} Torr]	[B] [1·10^{-4} Torr]
-1	0.8	x+0.8
0	1.0	x+1.0
+1	1.2	x+1.2

principle still be applied to these higher dimensional problems. However, as this procedure effectively projects the front onto the residual axis, the more complex structure of the pareto front is not accounted for. An extension of the algorithm to higher dimensions is currently not implemented. For this reason, the model selection in figure A.6 was done manually.

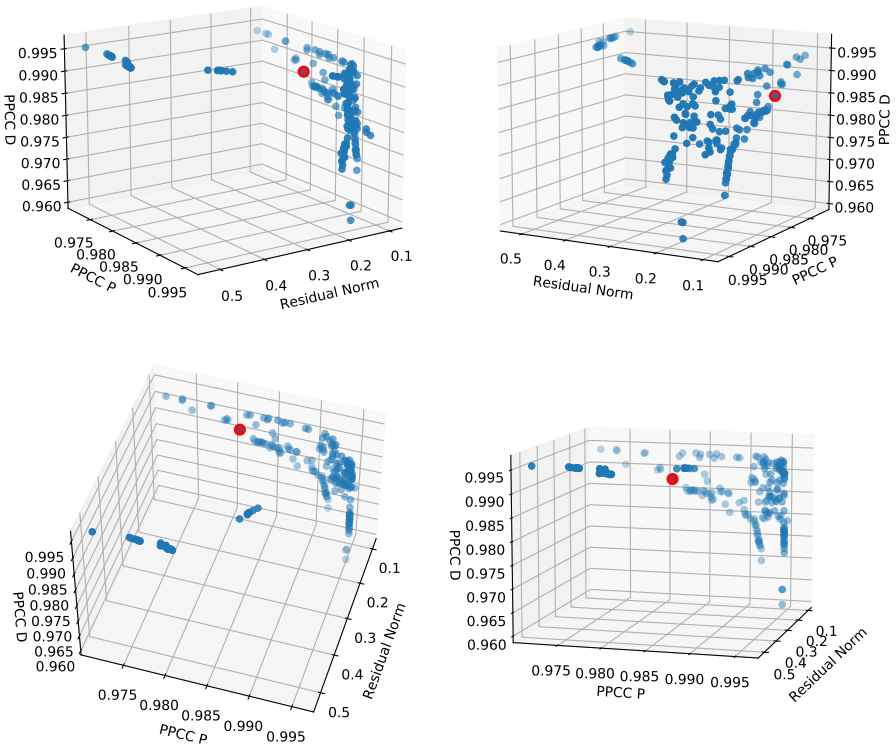


Figure A.6: Pareto front for the isothermal catalytic cycle model with substrate decomposition. The cost function for this problem is now 3D, as PPCC values for two response variables need to be optimized and the subplots show different angles of the same front. The optimization was performed on a population of 200 individuals. The red marker indicates the selected model.

The majority of solutions within this 3 dimensional pareto front are located at low residual values, with a semi-continuous domain of points at a residual norm of roughly 0.1. The typical ellbow shape of the front can be found in the plane of the two PPCC values. Ultimately, the solution, which was chosen from this front, is located at the edge of such a domain in the almost 2 dimensional sub front at low residual values.

To determine the temperature dependence of the two reaction pathways in the isothermal catalytic cycle with substrate decomposition, individual models for both product rates have been fit. The corresponding pareto fronts are given in figures A.8 and A.7. Both optimization

have been performed on the same dataset based on full factorial design with three factors ([A],[B], T).

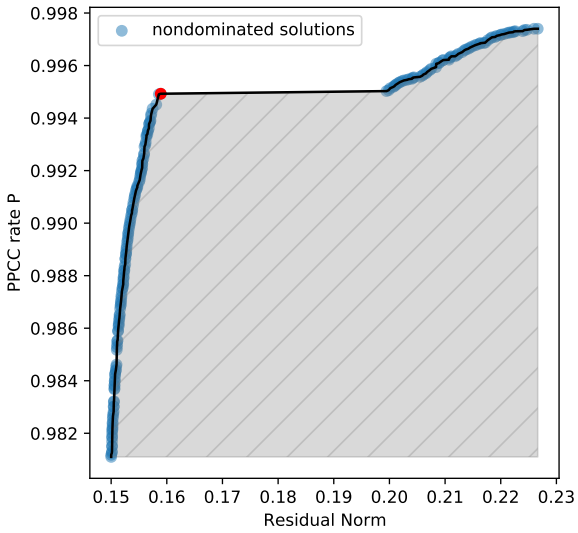


Figure A.7: Pareto front for the temperature dependent catalytic cycle model with substrate decomposition. The optimization was performed on a population of 200 individuals. The red marker indicates the selected model. The target quantity is the production rate of the main product P.

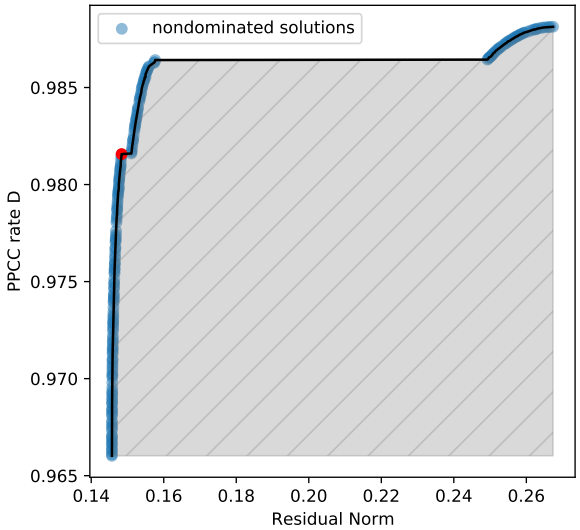


Figure A.8: Pareto front for the temperature dependent catalytic cycle model with substrate decomposition. The optimization was performed on a population of 200 individuals. The red marker indicates the selected model. The target quantity is the production rate of the side product D.

For these individual models the pareto front again becomes a 2 dimensional object. Therefore, the standard model selection could be applied. Both fronts show a similar behavior with a significant gap between low and high residual solutions.

A.2.4 CO Oxidation

Figure A.9 shows the pareto front for the maximum rate region in the CO oxidation example. Here we have an example for a pareto front, which makes the selection of clear optimal

Table A.5: Parameter scalings for the full factorial.

Level	$[O_2]$ [$1{\cdot}10^{-4}$ Torr]	[CO] [$1{\cdot}10^{-4}$ Torr]	$[CO_2]$ [$1{\cdot}10^{-4}$ Torr]	T [K]
-1	0.8	0.8	0.8	460
0	1.0	1.0	1.0	490
+1	1.2	1.2	1.2	520

solution quite difficult. The entire front is basically one continuous domain. This is often the case, if multiple (>2) LASSO terms contribute significantly the overall model, as in such cases there are various possible trade offs which result in very similar objective values. Still, in this case there are indeed some minor gaps, which allow for a model selection using the automated analysis implemented in our algorithm.

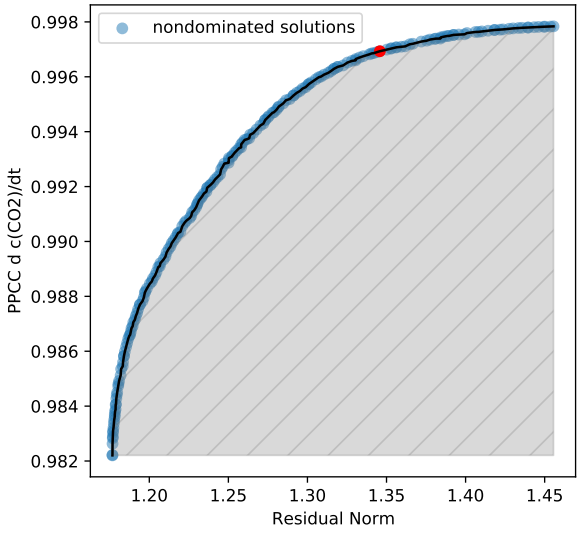


Figure A.9: Pareto front for the CO oxidation model. The optimization was performed on a population of 200 individuals. The red marker indicates the selected model. The target quantity is the production rate of CO_2.

This maximum region model was fit based on a 3-level full factorial design with scaling parameters as given in table A.5.

In Figure A.10 we show the predictions of the maximum region model along cuts through all 4 feature dimension (T, $[O_2]$, [CO], $[CO_2]$) compared to the data from the full factorial design.

Table A.6: Parameter scalings for the temperature scan designs.

Level	$[O_2]$ [$1{\cdot}10^{-4}$ Torr]	[CO] [$1{\cdot}10^{-4}$ Torr]	$[CO_2]$ [$1{\cdot}10^{-4}$ Torr]
-1	0.8	0.8	0.8
0	1.0	1.0	1.0
+1	1.2	1.2	1.2

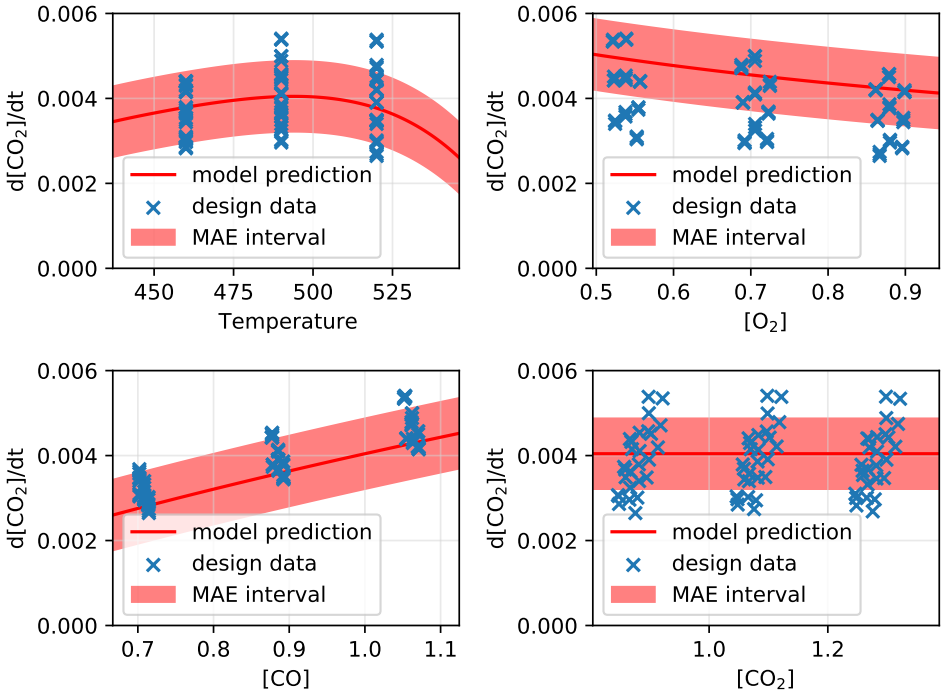


Figure A.10: Maximum region model predictions along cuts through all 4 feature dimensions.

A.2.5 Temperature Scan Designs

The temperature scan designs are basically composed of experimental designs, which are translated along the temperature axis. For every temperature value we evaluate a basis design for all $N-1$ remaining factors (partial pressures). In the case of the full factorial reference design, this basis design is simply a 3 factor (p_{CO}, p_{O_2}, p_{CO_2}) 3-level full factorial. For the screening designs, we used a Placket-Burman (PB) design [109] at every temperature. For every PB design the central point (0,0,0) was added to allow for a better estimation of nonlinear effects. The respective design matrices are given in Tables A.7 and A.8. Table A.6 contains the parameter settings for the partial pressures. The temperature was increased in steps of 20.0K for the full factorial design and steps of 23.5K for the screening design.

A.3 Toy Models for the Adaptive Design Algorithm

Here, the exact equations for the toy models discussed in chapter 3.3 are given. The transition functions $q(x_1, x_2)$ are arctangent type trigonometric functions centered around the transition boundaries. The error function $e(x_1, x_2)$ is used to scale the standard deviation of a gaussian noise term σ, which is drawn from a standard normal distribution. The function $e(x_1, x_2)$ itself is of gaussian shape and centered along the regime boundary, such that the

Table A.7: Temperature scan full factorial design with a total of 567 points for N=21.

T	$[O_2]$	$[CO]$	$[CO_2]$
T_1	-1	-1	-1
T_1	0	-1	-1
T_1	1	-1	-1
T_1	-1	0	-1
T_1	0	0	-1
T_1	1	0	-1
T_1	-1	1	-1
T_1	0	1	-1
T_1	1	1	-1
T_1	-1	-1	0
T_1	0	-1	0
T_1	1	-1	0
T_1	-1	0	0
T_1	0	0	0
T_1	1	0	0
T_1	-1	1	0
T_1	0	1	0
T_1	1	1	0
T_1	-1	-1	1
T_1	0	-1	1
T_1	1	-1	1
T_1	-1	0	1
T_1	0	0	1
T_1	1	0	1
T_1	-1	1	1
T_1	0	1	1
T_1	1	1	1
...	...	...	...
T_N	-1	-1	-1
T_N	0	-1	-1
T_N	1	-1	-1
T_N	-1	0	-1
T_N	0	0	-1
T_N	1	0	-1
T_N	-1	1	-1
...	...	...	...

Table A.8: Temperature scan screening design with a total of 90 datapoints for N=18.

T	$[O_2]$	[CO]	$[CO_2]$
T_1	-1	-1	1
T_1	1	-1	-1
T_1	-1	1	-1
T_1	1	1	1
T_1	0	0	0
...	...	...	...
T_N	-1	-1	1
T_N	1	-1	-1
T_N	-1	1	-1
T_N	1	1	1
T_N	0	0	0

magnitude of the noise term is larger the closer a point is to the boundary, resembling the expected behavior of a real reaction kinetic system.

A.3.1 Piecewise Constant Models with Linear Boundaries

Two Regimes with Linear Boundary
This model is shown in Figure 3.28 a).

$$
\begin{aligned}
f(x_1, x_2) &= -1 \cdot q(x_1, x_2) + 1 \cdot (1 - q(x_1, x_2)) + \sigma e(x_1, x_2) \cdot \\
q(x_1, x_2) &= \frac{arctan(20 \cdot (x_2 - s(x_1))) + \pi/2}{\pi} \\
e(x_1, x_2) &= 0.1 \cdot \exp\left(\frac{-(x_2 - s(x_1))^2}{4}\right) \\
s(x_1) &= 0.3 \cdot x_1 - 1.5
\end{aligned}
\tag{A.1}
$$

Three Regimes with Linear Boundaries without Intersection
This model is shown in Figure 3.28 b).

$$
\begin{aligned}
f(x_1,x_2) &= 5\cdot(1-q_1(x_1,x_2)) - 8\cdot q_2(x_1,x_2) + \sigma e(x_1,x_2)\cdot \\
q_1(x_1,x_2) &= \frac{arctan(20\cdot(x_2-s_1(x_1))) + \pi/2}{\pi} \\
q_2(x_1,x_2) &= \frac{arctan(20\cdot(x_2-s_2(x_1))) + \pi/2}{\pi} \\
e(x_1,x_2) &= 0.1\cdot\exp\left(\frac{-(x_2-s_1(x_1))^2}{4}\right) + 0.1\cdot\exp\left(\frac{-(x_2-s_2(x_1))^2}{4}\right) \\
s_1(x_1) &= -0.3\cdot x_1 - 2.5 \\
s_2(x_1) &= 0.1\cdot x_1 + 3.5
\end{aligned}
\tag{A.2}
$$

Three Regimes with Linear Boundaries and Triple Point
This model is shown in Figure 3.28 c).

$$
\begin{aligned}
f(x_1,x_2) &= 5\cdot(q_1(x_1,x_2)\cdot q_2(x_1,x_2)) - 8\cdot(1-q_1(x_1,x_2))\sigma e(x_1,x_2)\cdot \\
q_1(x_1,x_2) &= \frac{arctan(20\cdot(x_2-s_1(x_1))) + \pi/2}{\pi} \\
q_2(x_1,x_2) &= \frac{arctan(20\cdot(x_2-s_2(x_1))) + \pi/2}{\pi} \\
e(x_1,x_2) &= 0.1\cdot\exp\left(\frac{-(x_2-s_1(x_1))^2}{4}\right) + 0.1\cdot\exp\left(\frac{-(x_2-s_2(x_1))^2}{4}\right) \\
s_1(x_1) &= -0.3\cdot x_1 - 2.5 \\
s_2(x_1) &= 0.9\cdot x_1 + 1
\end{aligned}
\tag{A.3}
$$

A.3.2 Piecewise Nonconstant Models

Two Regimes with Linear Boundary
This model is shown in Figure 3.36 a).

$$
\begin{aligned}
f(x_1, x_2) &= x_1 \cdot q(x_1, x_2) + (1 - x_2^2) \cdot (1 - q(x_1, x_2)) + \sigma e(x_1, x_2) \cdot \\
q(x_1, x_2) &= \frac{arctan(20 \cdot (x_2 - s(x_1))) + \pi/2}{\pi} \\
e(x_1, x_2) &= 0.1 \cdot \exp\left(\frac{-(x_2 - s(x_1))^2}{4}\right) \\
s(x_1) &= 0.3 \cdot x_1 - 1.5
\end{aligned}
\tag{A.4}
$$

Two Regimes with Nonlinear Boundary
This model is shown in Figure 3.36 b).

$$
\begin{aligned}
f(x_1, x_2) &= (-0.2 \cdot (x_1 - 15)^2 - x_1) \cdot q(x_1, x_2) + (0.2 \cdot (x_2 - 15)^2 - 1.5 + x_2) \cdot (1 - q(x_1, x_2)) + \sigma e(x_1, x_2) \cdot \\
q(x_1, x_2) &= \frac{arctan(20 \cdot (x_2 - s(x_1))) + \pi/2}{\pi} \\
e(x_1, x_2) &= 0.1 \cdot \exp\left(\frac{-(x_2 - s(x_1))^2}{4}\right) \\
s(x_1) &= 0.03 \cdot (x_1 + 10)^2 - 5
\end{aligned}
\tag{A.5}
$$

A.4 Final Design Matrices of the Adaptive Design Algorithm

Here, the final adaptive designs for all test cases are given explicitly.

A.4.1 Piecewise Constant Models with Linear Boundaries

Two Regimes with Linear Boundary

Table A.9: Final adaptive design for the two regime piecewise constant system with linear boundary.

x_1	x_2
-1	-1
1	-1
-1	1
1	1
-0.435	-0
-0.597	-0.5
-0.1	-0.75
1	-0.016

Three Regimes with Linear Boundaries without Intersection

Table A.10: Final adaptive design for the three regime piecewise constant system with non intersecting linear boundaries.

x_1	x_2
-1	-1
1	-1
-1	1
1	1
-0.569	-0.785
1	0.108
0.696	0.554
1	-0.267
-0.304	-0.455
0.029	-0.227
-1	0.468
0.096	1
0.12	0.186
0.134	-1

Three Regimes with Linear Boundaries and Triple Point

Table A.11: Final adaptive design for the three regime piecewise constant system with intersecting linear boundaries.

x_1	x_2
-1	-1
1	-1
-1	1
1	1
-0.065	-0
0.882	-0.272
1	0.5
0.353	-0.636
-0.152	0.75
-1	-0.538
-1	0.582
1	-0.389
-0.334	0.094
0.18	-1
-0.715	-0.118

A.4.2 Piecewise Nonconstant Models

Two Regimes with Linear Boundary

Table A.12: Final adaptive design for the two regime piecewise nonconstant system with linear boundary.

x_1	x_2
-1	-1
1	-1
-1	1
1	1
0.698	-0
0.213	-0.5
-1	-0.027
-1	0.483
1	0.526
-1	-0.461

Two Regimes with Nonlinear Boundary

Table A.13: Final adaptive design for the two regime piecewise nonconstant system with nonlinear boundary.

x_1	x_2
-1	-1
1	-1
-1	1
1	1
0	0
0.29	-0.5
-1	0.162
1	-0.191
1	1
-1	1
1	-1
-0.846	-0.029
0.705	0.048
-0.417	-0.271
-0.172	-0.633
-1	-0.653
1	0.618
0.515	1
0.961	0.49
-1	-0.248
1	0.878

A.4.3 Isothermal CO Oxidation over RuO_2

Table A.14: Final adaptive design for the isothermal CO oxidation reaction over RuO_2.

$\log(p_{CO})$	$\log(p_{O_2})$
-1	-1
1	-1
-1	1
1	1
0	0
0.6	0.321
0.441	1
-0.014	-0.924
-0.202	1
0.516	1
-0.397	-1
0.129	-1

Table A.15: Final adaptive design for the isothermal CO oxidation reaction over RuO_2 using a batch size of 3.

$\log(p_{CO})$	$\log(p_{O_2})$
-1	-1
1	-1
-1	1
1	1
0	0
0.6	0.915
0.85	0.733
-0.101	1
0.916	-0.147
-0.245	1
-0.258	-1
0.075	-1
-1	-0.196

A.4.4 Temperature Dependent CO Oxidation over RuO_2

Table A.16: Final adaptive design for the CO oxidation reaction over RuO_2 with varying temperature.

$\log(p_{CO})$	$\log(p_{O_2})$	T
-1	-1	1
1	-1	-1
-1	1	-1
1	1	1
0	0	0
0.285	0.026	0.861
0.219	-0.45	0.519
0.988	-0.825	0.143
-0.609	-0.232	0.088
-0.516	0.173	0.664
0.222	0.022	-0.637
-1	-1	-1
1	0.003	-1
-1	1	1
1	-1	1
-0.152	-1	-1
1	0.945	-1
-0.231	1	-1
-0.318	1	0.232
-0.006	1	1
-1	-0.178	1
-1	-0.2	-1
-1	1	-0.041

Bibliography

[1] F. Felsen, K. Reuter, and C. Scheurer. "A model-free sparse approximation approach to robust formal reaction kinetics". In: *Chem. Eng. J.* 433 (2022), p. 134121. ISSN: 1385-8947. https://doi.org/10.1016/j.cej.2021.134121.

[2] G. Box and N. Draper. *Response Surfaces, Mixtures, and Ridge Analyses*. Wiley Series in Probability and Statistics. Wiley, 2007.

[3] K. Reuter and M. Scheffler. "First-Principles Atomistic Thermodynamics for Oxidation Catalysis: Surface Phase Diagrams and Catalytically Interesting Regions". In: *Phys. Rev. Lett.* 90 (2003), p. 046103. https://doi.org/doi/10.1103/PhysRevLett.90.046103.

[4] M. Andersen, C. Panosetti, and K. Reuter. "A Practical Guide to Surface Kinetic Monte Carlo Simulations". In: *Front. Chem.* 7 (2019), p. 202. DOI: 10.3389/fchem.2019.00202.

[5] P. Stoltze. "Microkinetic simulation of catalytic reactions". In: *Prog. Surf. Sci.* 65.3 (2000), pp. 65 –150. DOI: 10.1016/S0079-6816(00)00019-8.

[6] F. Santosa and W. W. Symes. "Linear Inversion of Band-Limited Reflection Seismograms". In: *SIAM Journal on Scientific and Statistical Computing* 7.4 (1986), pp. 1307–1330. DOI: 10.1137/0907087. URL: https://doi.org/10.1137/0907087.

[7] J. J. Filliben. "The Probability Plot Correlation Coefficient Test for Normality". In: *Technometrics* 17.1 (1975), pp. 111–117. DOI: 10.1080/00401706.1975.10489279.

[8] K. Deb et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II". In: *IEEE Transactions on Evolutionary Computation* 6.2 (2002), pp. 182–197. DOI: 10.1109/4235.996017.

[9] C.-C. Chang and C.-J. Lin. "LIBSVM: a library for support vector machines". In: *ACM Trans. Intell. Syst. Technol.* 2 (2011), p. 1.

[10] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. Springer New York, 2009. ISBN: 9780387781655. URL: https://books.google.de/books?id=wj5OyydZ5bkC.

[11] V. Fedorov. *Theory of Optimal Experiments*. Probability and Mathematical Statistics. Academic Press. URL: https://books.google.de/books?id=PpUIwgEACAAJ.

[12] B. Temel et al. "Does phenomenological kinetics provide an adequate description of heterogeneous catalytic reactions?" In: *J. Chem. Phys.* 126 (2007), p. 204711.

[13] J. H. van't Hoff, E. Cohen, and T. Ewan. *Studies in chemical dynamics*. Williams & Norgate, 1896.

[14] O. Wolkenhauer et al. "Modelling reaction kinetics inside cells". In: *Essays Biochem.* 45 (2008), p. 41. https://doi.org/10.1042/bse0450041.

[15] G. Schreiber. "Kinetic studies of protein–protein interactions". In: *Curr. Opin. Struct. Biol.* 12 (2002), p. 41. https://doi.org/10.1016/S0959-440X(02)00287-7.

[16] W. Cleland. "Enzyme kinetics". In: *Annu. Rev. Biochem.* 36 (1967), p. 77.

[17] A. Cornish-Bowden. "One hundred years of Michaelis–Menten kinetics". In: *Perspect. Sci.* 4 (2015), p. 3. https://doi.org/10.1016/j.pisc.2014.12.002.

[18] G. R. Wittreich, K. Alexopoulos, and D. G. Vlachos. "Microkinetic modeling of surface catalysis". In: *Handbook of Materials Modeling: Applications: Current and Emerging Materials* (2020), p. 1377.

[19] R. Chaudhari, A Seayad, and S Jayasree. "Kinetic modeling of homogeneous catalytic processes". In: *Catal. Today* 66 (2001), p. 371. https://doi.org/10.1016/S0920-5861(00)00633-7.

[20] M. Besora and F. Maseras. "Microkinetic modeling in homogeneous catalysis". In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 8 (2018), e1372.

[21] C. Hill and T. Root. *Introduction to Chemical Engineering Kinetics and Reactor Design*. Wiley, 2014. ISBN: 9781118797839.

[22] G. A. Olah. "Beyond oil and gas: the methanol economy". In: *Angew. Chem. Int. Ed.* 44 (2005), p. 2636.

[23] K. Reuter. "Ab initio thermodynamics and first-principles microkinetics for surface catalysis". In: *Catal. Lett.* 146 (2016), p. 541. https://doi.org/10.1007/s10562-015-1684-3.

[24] B. W. J. Chen, L. Xu, and M. Mavrikakis. "Computational Methods in Heterogeneous Catalysis". In: *Chem. Rev.* 121 (2021), p. 1007. https://doi.org/10.1021/acs.chemrev.0c01060.

[25] A. H. Motagamwala, M. R. Ball, and J. A. Dumesic. "Microkinetic analysis and scaling relations for catalyst design". In: *Annu. Rev. Chem. Biomol. Eng.* 9 (2018), p. 413.

[26] A. H. Motagamwala and J. A. Dumesic. "Microkinetic Modeling: A Tool for Rational Catalyst Design". In: *Chem. Rev.* 121 (2021), p. 1049. https://doi.org/10.1021/acs.chemrev.0c00394.

[27] A. Bruix et al. "First-principles-based multiscale modelling of heterogeneous catalysis". In: *Nature Catal.* 2 (2019), p. 659.

[28] J. E. Sutton et al. "Electrons to reactors multiscale modeling: catalytic CO oxidation over RuO2". In: *ACS Catal.* 8 (2018), p. 5002. https://doi.org/10.1021/acscatal.8b00713.

[29] C. T. Campbell. "Future directions and industrial perspectives micro-and macro-kinetics: their relationship in heterogeneous catalysis". In: *Top. Catal.* 1 (1994), p. 353.

[30] M. Lobera et al. "Transient kinetic modelling of propane dehydrogenation over a Pt–Sn–K/Al2O3 catalyst". In: *Appl. Catal. A: Gen.* 349 (2008), p. 156. https://doi.org/10.1016/j.apcata.2008.07.025.

[31] L. C. Grabow et al. "Mechanism of the water gas shift reaction on Pt: First principles, experiments, and microkinetic modeling". In: *J. Phys. Chem. C* 112 (2008), p. 4608. https://doi.org/10.1021/jp7099702.

[32] N. Park et al. "Kinetic modeling of methanol synthesis over commercial catalysts based on three-site adsorption". In: *Fuel Process. Technol.* 125 (2014), p. 139. https://doi.org/10.1016/j.fuproc.2014.03.041.

[33] G. Box, J. Hunter, and W. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley, 2005.

[34] V. Zhdanov and B. Kasemo. "Kinetic phase transitions in simple reactions on solid surfaces". In: *Surf. Sci. Rep.* 20 (1994), p. 113. https://doi.org/10.1016/0167-5729(94)90009-4.

[35] N. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library. Elsevier Science, 1992. ISBN: 9780080571386. URL: https://books.google.de/books?id=3e7XbMoJzmoC.

[36] C. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer complexity. Springer, 2004. ISBN: 9783540208822. URL: https://books.google.de/books?id=wLm7QgAACAAJ.

[37] M. Feinberg. *Foundations of Chemical Reaction Network Theory*. Springer International Publishing, 2019.

[38] A. Hellman and K Honkala. "Including lateral interactions into microkinetic models of catalytic reactions". In: *The Journal of chemical physics* 127.19 (2007), p. 194704.

[39] B. W. J. Chen, L. Xu, and M. Mavrikakis. "Computational Methods in Heterogeneous Catalysis". In: *Chem. Rev.* 121.2 (2021), pp. 1007–1048. DOI: 10.1021/acs.chemrev.0c01060.

[40] L. Fahrmeir et al. *Statistik: Der Weg zur Datenanalyse*. Springer Berlin Heidelberg, 2016.

[41] T. Cleff. *Deskriptive Statistik und Explorative Datenanalyse: Eine computergestützte Einführung mit Excel, SPSS und STATA*. Gabler Verlag, 2015.

[42] R. Christensen. *Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization*. Springer, 2001.

[43] P. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, 1997.

[44] G. James et al. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer New York, 2014.

[45] R. Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 00359246.

[46] J. Friedman, T. Hastie, and R. Tibshirani. "Regularization paths for generalized linear models via coordinate descent". In: *J. Stat. Softw.* 33 (2010), p. 1.

[47] H. Zou and T. Hastie. "Regularization and variable selection via the elastic net". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (2005), pp. 301–320. DOI: https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[48] D. Montgomery. *Design and Analysis of Experiments*. Wiley, 2017.

[49] K. Siebertz, D. van Bebber, and T. Hochkirchen. *Statistische Versuchsplanung: Design of Experiments (DoE)*. Springer Berlin Heidelberg, 2010.

[50] Z. Lazic. *Design of Experiments in Chemical Engineering: A Practical Guide*. Wiley, 2004. ISBN: 9783527311422. URL: https://books.google.de/books?id=79JTAAAAMAAJ.

[51] S. Shina. *Industrial Design of Experiments: A Case Study Approach for Design and Process Optimization*. Springer International Publishing, 2022. ISBN: 9783030862671. URL: https://books.google.de/books?id=MJtXEAAAQBAJ.

[52] G. E. P. Box and D. W. Behnken. "Some New Three Level Designs for the Study of Quantitative Variables". In: *Technometrics* 2.4 (1960), pp. 455–475. DOI: 10.1080/00401706.1960.10489912.

[53] S. Silvey and C. Chatfield. *Optimal Design: An Introduction to the Theory for Parameter Estimation*. Ettore Majorana International Science Series Bd. 10. Springer Netherlands, 1980. ISBN: 9780412229107. URL: https://books.google.de/books?id=uXGmAAAAIAAJ.

[54] V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments*. Lecture Notes in Statistics. Springer New York, 2012. ISBN: 9781461207030. URL: https://books.google.de/books?id=6pgRBwAAQBAJ.

[55] J. Tukey. *Exploratory Data Analysis*. Addison-Wesley series in behavioral science Bd. 2. Addison-Wesley Publishing Company, 1977. ISBN: 9780201076165.

[56] F. Mosteller and J. Tukey. *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley series in behavioral science. Addison-Wesley Publishing Company, 1977. ISBN: 9780201048544.

[57] G. E. P. Box and D. R. Cox. "An Analysis of Transformations". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 26.2 (1964), pp. 211–243. DOI: 10.1111/j.2517-6161.1964.tb00553.x.

[58] I. Yeo and R. A. Johnson. "A new family of power transformations to improve normality or symmetry". In: *Biometrika* 87.4 (Dec. 2000), pp. 954–959. ISSN: 0006-3444. DOI: 10.1093/biomet/87.4.954.

[59] B. Panigrahi, Y. Shi, and M. Lim. *Handbook of Swarm Intelligence: Concepts, Principles and Applications*. Adaptation, Learning, and Optimization. Springer Berlin Heidelberg, 2011. ISBN: 9783642173905. URL: https://books.google.de/books?id=LYsbMvMRvckC.

[60] Y. Zhang, S. Wang, and G. Ji. "A comprehensive survey on particle swarm optimization algorithm and its applications". In: *Mathematical problems in engineering* 2015 (2015).

[61] E. Polak. *Optimization: Algorithms and Consistent Approximations*. Applied Mathematical Sciences. Springer New York, 1997.

[62] Y. Collette and P. Siarry. *Multiobjective Optimization: Principles and Case Studies*. Decision Engineering. Springer Berlin Heidelberg, 2013. ISBN: 9783662088838. URL: https://books.google.de/books?id=dIrqCAAAQBAJ.

[63] J. Branke et al., eds. *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Berlin, Heidelberg: Springer-Verlag, 2008. ISBN: 9783540889076.

[64] J. Vasconcelos et al. "Improvements in genetic algorithms". In: *IEEE Trans. Magn.* 37 (2001), p. 3414.

[65] E. Zitzler, D. Brockhoff, and L. Thiele. "The Hypervolume Indicator Revisited: On the Design of Pareto-compliant Indicators Via Weighted Integration". In: *Evolutionary Multi-Criterion Optimization*. Springer Berlin Heidelberg, 2007, p. 862.

[66] J. Blank and K. Deb. "Pymoo: Multi-Objective Optimization in Python". In: *IEEE Access* 8 (2020).

[67] J. Dick, F. Y. Kuo, and I. H. Sloan. "High-dimensional integration: The quasi-Monte Carlo way". In: *Acta Numerica* 22 (2013), 133–288. DOI: 10.1017/S0962492913000044.

[68] J. H. Halton. "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals". In: *Numer. Math.* 2 (1960), p. 84.

[69] I. M. Sobol'. "On the distribution of points in a cube and the approximate evaluation of integrals". In: *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* 7.4 (1967), p. 784.

[70] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009. ISBN: 9780387848846. URL: https://books.google.de/books?id=eBSgoAEACAAJ.

[71] Q. Du, V. Faber, and M. Gunzburger. "Centroidal Voronoi Tessellations: Applications and Algorithms". In: *SIAM Review* 41.4 (1999), pp. 637–676. DOI: 10.1137/S0036144599352836. eprint: https://doi.org/10.1137/S0036144599352836. URL: https://doi.org/10.1137/S0036144599352836.

[72] Q. Du, M. Emelianenko, and L. Ju. "Convergence of the Lloyd Algorithm for Computing Centroidal Voronoi Tessellations". In: *SIAM Journal on Numerical Analysis* 44.1 (2006), pp. 102–119. DOI: 10.1137/040617364. eprint: https://doi.org/10.1137/040617364. URL: https://doi.org/10.1137/040617364.

[73] J. C. Chen, E. A. Rubin, and G. J. Cornwall. *Data Science for Public Policy*. Cham: Springer International Publishing, 2021, pp. 217–235. ISBN: 978-3-030-71352-2. DOI: 10.1007/978-3-030-71352-2_11. URL: https://doi.org/10.1007/978-3-030-71352-2_11.

[74] V. Satopaa et al. "Finding a" kneedle" in a haystack: Detecting knee points in system behavior". In: *2011 31st international conference on distributed computing systems workshops*. IEEE. 2011, pp. 166–171.

[75] C. Bishop. *Pattern Recognition and Machine Learning: All "just the Facts 101" Material*. Information science and statistics. Springer (India) Private Limited, 2013. ISBN: 9788132209065. URL: https://books.google.de/books?id=HL4HrgEACAAJ.

[76] J. Platt et al. "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods". In: *Advances in large margin classifiers* 10.3 (1999), pp. 61–74.

[77] T.-F. Wu, C.-J. Lin, and R. C. Weng. "Probability estimates for multi-class classification by pairwise coupling". In: *Journal of Machine Learning Research* 5.Aug (2004), pp. 975–1005.

[78] M. Bodenstein and S. C. Lind. "Geschwindigkeit der Bildung des Bromwasserstoffs aus seinen Elementen". In: *Z. Phys. Chem.* ().

[79] K. F. Herzfeld. "Zur Theorie der Reaktionsgeschwindigkeiten in Gasen". In: *Ann. Phys.* 364.15 (1919), pp. 635–667. DOI: 10.1002/andp.19193641504.

[80] A. Levy. "The High-Temperature Kinetics of the Hydrogen-Bromine Reaction". In: *J. Phys. Chem.* 62.5 (1958), pp. 570–574. DOI: 10.1021/j150563a013.

[81] S. D. Cooley and R. C. Anderson. "Flame Propagation Studies Using the Hydrogen-Bromine Reaction." In: *Ind. Eng. Chem.* 44.6 (1952), pp. 1402–1406. DOI: 10.1021/ie50510a058.

[82] R. Pollice. "A General Fitting Function to Estimate Apparent Reaction Orders of Kinetic Profiles". In: *ChemRxiv* (Mar. 2019). DOI: 10.26434/chemrxiv.7885760.v1.

[83] T. Engel and G. Ertl. "Elementary Steps in the Catalytic Oxidation of Carbon Monoxide on Platinum Metals". In: ed. by D. Eley, H. Pines, and P. B. Weez. Vol. 28. Adv. Catal. Academic Press, 1979. DOI: 10.1016/S0360-0564(08)60133-9.

[84] R. Imbihl and G. Ertl. "Oscillatory Kinetics in Heterogeneous Catalysis". In: *Chem. Rev.* 95.3 (1995), pp. 697–733. DOI: 10.1021/cr00035a012.

[85] J. Wang et al. "The Kinetics of CO Oxidation on RuO2(110): Bridging the Pressure Gap". In: *J. Phys. Chem. B* 106 (2002), p. 3422. https://doi.org/10.1021/jp014109k.

[86] S. Matera, H. Meskine, and K. Reuter. "Adlayer inhomogeneity without lateral interactions: Rationalizing correlation effects in CO oxidation at RuO2(110) with first-principles kinetic Monte Carlo". In: *J. Chem. Phys.* 134 (2011), p. 064713. https://doi.org/10.1063/1.3553258.

[87] J. E. Sutton et al. "Electrons to Reactors Multiscale Modeling: Catalytic CO Oxidation over RuO2". In: *ACS Catal.* 8 (2018), p. 5002. https://doi.org/10.1021/acscatal.8b00713.

[88] S. Wendt, M. Knapp, and H. Over. "The Role of Weakly Bound On-Top Oxygen in the Catalytic CO Oxidation Reaction over RuO2(110)". In: *J. Am. Chem. Soc.* 126 (2004), p. 1537. https://doi.org/10.1021/ja0364423.

[89] A. Farkas, F. Hess, and H. Over. "Experiment-Based Kinetic Monte Carlo Simulations: CO Oxidation over RuO2(110)". In: *J. Phys. Chem. C.* 116 (2012), p. 581. https://doi.org/10.1021/jp204703p.

[90] F. A. Lindemann et al. "Discussion on 'the radiation theory of chemical action'". In: *Trans. Faraday Soc.* 17 (0 1922), pp. 598–606. DOI: 10.1039/TF9221700598.

[91] H. Gossler et al. "CaRMeN: An Improved Computer-Aided Method for Developing Catalytic Reaction Mechanisms". In: *Catalysts* 9.3 (2019), p. 227.

[92] D. G. Blackmond. "Reaction Progress Kinetic Analysis: A Powerful Methodology for Mechanistic Studies of Complex Catalytic Reactions". In: *Angew. Chem. Int. Ed.* 44.28 (2005), pp. 4302–4320. DOI: https://doi.org/10.1002/anie.200462544.

[93] J. Burés. "What is the Order of a Reaction?" In: *Top. Catal.* 60 (2017), pp. 631–633. DOI: 10.1007/s11244-017-0735-y.

[94] J. Burés. "A Simple Graphical Method to Determine the Order in Catalyst". In: *Angew. Chem. Int. Ed.* 55 (2016), pp. 2028 –2031. DOI: 10.1002/anie.201508983.

[95] J. Burés. "Variable Time Normalization Analysis: General Graphical Elucidation of Reaction Orders from Concentration Profiles." In: *Angew. Chem. Int. Ed.* 55 52 (2016), pp. 16084–16087. DOI: 10.1002/anie.201609757.

[96] M. T. G.E. Garrett. "A Nonlinear Ordinary Differential Equation for Generating Graphical Rate Equations from Concentration Versus Time Data." In: *Top. Catal.* 60 (2017), pp. 554–563. DOI: 10.1007/s11244-017-0739-7.

[97] K. J. Laidler. "A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC Recommendations 1996)". In: *Pure Appl. Chem.* 68.1 (1996), pp. 149–192. DOI: 10.1351/pac199668010149.

[98] J. Blank and K. Deb. "Pymoo: Multi-Objective Optimization in Python". In: *IEEE Access* 8 (2020), pp. 89497–89509.

[99] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12 (2011), pp. 2825–2830.

[100] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nat. Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[101] W. Schiesser. *The Numerical Method of Lines: Integration of Partial Differential Equations*. Academic Press, 1991.

[102] G. E. P. Box, W. H. Hunter, S. Hunter, et al. *Statistics for experimenters*. Vol. 664. John Wiley and sons New York, 1978.

[103] E. M. Cordos, S. R. Crouch, and H. V. Malmstadt. "Automatic digital readout system for reaction-rate methods". In: *Anal. Chem.* 40.12 (1968), pp. 1812–1818. DOI: 10.1021/ac60268a018.

[104] L. L. Joffrion et al. "The Effect of Experimental Measurement Uncertainties on Rate Modeling And Reactor Simulation". In: *Chem. Eng. Commun.* 76.1 (1989), pp. 41–51. DOI: 10.1080/00986448908940317.

[105] B. R. Gaines, J. Kim, and H. Zhou. "Algorithms for Fitting the Constrained Lasso". In: *J. Comput. Graph. Stat.* 27.4 (2018), pp. 861–871. DOI: 10.1080/10618600.2018.1473777.

[106] R. Kissel-Osterrieder et al. "Experimental and theoretical investigation of CO oxidation on platinum: Bridging the pressure and materials gap". In: *Proc. Combust. Inst.* 28.1 (2000), pp. 1341–1348. DOI: 10.1016/S0082-0784(00)80348-3.

[107] Y. Y. Yeo, L. Vattuone, and D. A. King. "Calorimetric heats for CO and oxygen adsorption and for the catalytic CO oxidation reaction on Pt111". In: *J. Chem. Phys.* 106.1 (1997), pp. 392–401. DOI: 10.1063/1.473203.

[108] G. T. K. K. Gunasooriya and M. Saeys. "CO Adsorption on Pt(111): From Isolated Molecules to Ordered High-Coverage Structures". In: *ACS Catal.* 8.11 (2018), pp. 10225–10233. DOI: 10.1021/acscatal.8b02371.

[109] R. L. Placket and J. P. Burman. "The Design of Optimum Multifactorial Experiments". In: *Biometrika* 33 (1946), p. 305. https://doi.org/10.1093/biomet/33.4.305.

[110] I. Surowiec et al. In: ().

[111] R. Harman, L. Filová, and P. Richtárik. "A Randomized Exchange Algorithm for Computing Optimal Approximate Designs of Experiments". In: *J. Am. Stat. Assoc.* 115.529 (2020), pp. 348–361. DOI: 10.1080/01621459.2018.1546588.

[112] J. M. Berty. "Testing Commercial Catalysts in Recycle Reactors". In: *Catal. Rev. Sci. Eng.* 20 (1979), p. 75. https://doi.org/10.1080/03602457908065106.

[113] L. Caldwell. "An improved internal gas recirculation reactor for catalytic studies". In: *Appl. Catal.* 8 (1983), p. 199. https://doi.org/10.1016/0166-9834(83)80080-3.

[114] E. M. Calverley et al. "A small, well-mixed reactor for high throughput study of commercial catalyst pills". In: *Chem. Eng. Sci.* 151 (2016), p. 130. https://doi.org/10.1016/j.ces.2016.05.011.

[115] R. Dickman. "Kinetic phase transitions in a surface-reaction model: Mean-field theory". In: *Phys. Rev. A* 34 (1986), p. 4246. https://doi.org/10.1103/PhysRevA.34.4246.

[116] R. M. Ziff, E. Gulari, and Y. Barshad. "Kinetic Phase Transitions in an Irreversible Surface-Reaction Model". In: *Phys. Rev. Lett.* 56 (1986), p. 2553. https://doi.org/10.1103/PhysRevLett.56.2553.

[117] S. Wehner et al. "Kinetic phase transitions in the reaction CO+O→CO2 on Ir(111) surfaces". In: *J. Chem. Phys.* 119 (2003), p. 6831. https://doi.org/10.1063/1.1603212.

[118] M. Ankerst et al. "OPTICS: Ordering points to identify the clustering structure". In: *ACM SIGMOD Rec.* 28 (1999), p. 49.

[119] M. Z. Rodriguez et al. "Clustering algorithms: A comparative approach". In: *PloS one* 14 (2019), e0210236.

[120] H. Meskine et al. "Examination of the concept of degree of rate control by first-principles kinetic Monte Carlo simulations". In: *Surf. Sci.* 603 (2009), p. 1724.

[121] S. Silvey. *Optimal Design: An Introduction to the Theory for Parameter Estimation*. Ettore Majorana International Science Series. Springer Netherlands, 2013. ISBN: 9789400959125. URL: https://books.google.de/books?id=c6H6CAAAQBAJ.

[122] A. Parkinson, C. Sorensen, and N. Pourhassan. "A General Approach for Robust Optimal Design". In: *J. Mech. Des.* 115 (1993), p. 74. https://doi.org/10.1115/1.2919328.

[123] J. Mockus. *Bayesian Approach to Global Optimization: Theory and Applications*. Mathematics and its Applications. Springer Netherlands, 2012. ISBN: 9789400909090. URL: https://books.google.de/books?id=VuKoCAAAQBAJ.

[124] J. Peng et al. "Navigating multimetallic catalyst space with Bayesian optimization". In: *Joule* 5 (2021), p. 3069. https://doi.org/10.1016/j.joule.2021.11.011.

[125] P. Schlexer Lamoureux et al. "Machine Learning for Computational Heterogeneous Catalysis". In: *ChemCatChem* 11 (2019), p. 3581. https://doi.org/10.1002/cctc.201900595.

[126] J. Vandermause et al. "Active learning of reactive Bayesian force fields: Application to heterogeneous hydrogen-platinum catalysis dynamics". In: *arXiv* (2021). https://doi.org/10.48550/ARXIV.2106.01949.

[127] Z. Li, S. Wang, and H. Xin. "Toward artificial intelligence in catalysis". In: *Nature Catalysis* 1 (2018), p. 641.

[128] W. Yang, T. T. Fidelis, and W.-H. Sun. "Machine Learning in Catalysis, From Proposal to Practicing". In: *ACS Omega* 5 (2020), p. 83. https://doi.org/10.1021/acsomega.9b03673.

[129] P. van Helden, J. A. van den Berg, and R. L. J. Coetzer. "A statistical approach to microkinetic analysis". In: *Ind. Eng. Chem. Res.* 51.19 (2012), p. 6631.

[130] M. Bracconi and M. Maestri. "Training set design for machine learning techniques applied to the approximation of computationally intensive first-principles kinetic models". In: *Chem. Eng. J.* 400 (2020), p. 125469. https://doi.org/10.1016/j.cej.2020.125469.

[131] E. Brochu, M. W. Hoffman, and N. de Freitas. "Portfolio allocation for Bayesian optimization". In: *arXiv* (2010).

[132] J. Shi and T. Choi. *Gaussian Process Regression Analysis for Functional Data*. Taylor & Francis, 2011. ISBN: 9781439837733. URL: https://books.google.de/books?id=DkgdN6dRAicC.

[133] *Particle swarm optimization (PSO) with constraint support*. https://pythonhosted.org/pyswarm/. Accessed: 2022-05-01.

[134] Y. Eldar et al. "The farthest point strategy for progressive image sampling". In: *IEEE Trans. Image Process.* 6 (1997), p. 1305.

[135] A. P. Bartók et al. "Machine learning unifies the modeling of materials and molecules". In: *Sci. Adv.* 3 (2017), e1701816. https://doi.org/10.1126/sciadv.1701816.

[136] Z. Marzeh, M. Tahmasbi, and N. Mirehi. "Algorithm for finding the largest inscribed rectangle in polygon". In: *J. Algorithm Comput.* 51 (2019), p. 29.

[137] Y.-E. Li, D. Willcox, and R. D. Gonzalez. "Determination of rate constants by the frequency response method: CO on Pt/SiO2". In: *AIChE Journal* 35.3 (1989), p. 423. https://doi.org/10.1002/aic.690350310.

[138] C. Urmès et al. "Microkinetic Modeling of Acetylene Hydrogenation Under Periodic Reactor Operation". In: *ChemCatChem* 14 (2022), e202101826. https://doi.org/10.1002/cctc.202101826.

[139] A. Lie, J. Hoebink, and G. Marin. "The effects of oscillatory feeding of CO and O2 on the performance of a monolithic catalytic converter of automobile exhaust gas: a modelling study". In: *Chem. Eng. J. Biochem. Eng. J.* 53 (1993), p. 47. https://doi.org/10.1016/0923-0467(93)80006-I.

[140] W. Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.