A SIMULTANEOUS OPTIMIZATION APPROACH TO OPTIMAL EXPERIMENTAL DESIGN

vorgelegt von
Dipl.-Ing. Duc Minh Hoang
aus Ho-Chi-Minh-Stadt, Vietnam

von der Fakultät III - Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademisches Grades

Doktor der Ingenieurwissenschaften
- Dr. Ing. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr.-Ing. Matthias Kraume
Gutachterin:	Prof. Martha A. Grover
Gutachter:	Prof. Dr.-Ing. habil. Prof. Dr. h.c. Günter Wozny
Gutachter:	Prof. Dr.-Ing. Harvey Arellano-Garcia

Tag der wissenschaftlichen Aussprache: 19. Mai 2014

Berlin 2014
D 83

ACKNOWLEDGEMENT

This thesis was written during my work as a research assistant at the DBTA department (Process Dynamics and Operation) of the TU-Berlin. It is also a part of the project A3 of the Collaborative Research Center InPROMPT SFB/Transregio63 'Integrated Chemical Processes in Liquid Multiphase Systems'.

My special thank is dedicated to the head of the department Prof. Dr.-Ing. habil. Prof. h.c. Dr. h.c. Günter Wozny for his supervision, support, valuable suggestions and comments during my entire research time.

I wish to express my deepest appreciation to the project leader of InPROMPT Prof. Dr.-Ing. Matthias Kraume who has kindly agreed to be the chairman, and Prof. Martha A. Grover and Prof. Dr.-Ing. Harvey Arellano-Garcia who have kindly agreed to be the second and third reviewer of the doctoral committee.

My sincere gratitude goes to Prof. Dr.-Ing. Harvey Arellano-Garcia for his generous supports since my first study years until today. I also would like to express my deepest thanks to all colleagues and friends, especially Dr. Tilman Barz, Dr. Shankui Song, Mr. Robert Kraus and Mrs. Sandra Fillinger for their contributions, supports and valuable suggestions during my time at the department.

I wish to extend my deepest thanks to my family who has been continuously supporting and encouraging me throughout my education, and especially my father, whose wise advises always guide me through my life.

Finally, I would like to acknowledge the German Research Foundation (DFG-Deutsche Forschungsgesellschaft) for the financial support.

Duc Minh Hoang Berlin, March 2014

ABSTRACT

Modelbasierte optimale Versuchsplanung hat in den letzten beiden Jahrzehnten immer mehr an Bedeutung gewonnen, vor allem bei verfahrenstechnischen, bio-chemischen sowie chemischen Anwendungen. Durch geplante Kontrolltrajektorien können Experimente so optimiert werden, dass maximaler Informationsgehalt bei minimalem Experimentaufwand gewonnen werden kann. Das Potential, Zeit und Ressourcen dadurch zu sparen, ist in den meisten Anwendungsfällen enorm. Mit dem simultanen Optimierungsansatz wird eine effiziente Methode vorgestellt, die erfolgreich auf die Problemstellung der optimalen Versuchsplanung angewendet wurde. Der Hauptbeitrag dieser Arbeit ist die mathematische Herleitung der optimalen Versuchsplanung als simultanes Optimierungsproblem, dessen Kernpunkt in der totalen Diskretisierung des Systemmodels und der Sensitivitätsgleichungen liegt. Damit wird das ursprüngliche Optimalsteuerungsproblem in ein NLP Problem umgewandelt, das effizienter und robuster mit globalen Optimierungsalgorithmen gelöst wird. Mit dem vorgestellten Ansatz können alle Anforderungen an ein fortgeschrittenes optimales Versuchs-planungsproblem adressiert werden. Dieses beinhaltet die flexible Verwendung von Kontrolltrajektorien unterschiedlicher Ordnungen, die Optimierung von Anfangs-bedingungen, striktes Einhalten der Steuerungs- und Prozessbeschränkungen, die Einschränkung an Probenentnahmen und Experimentdauer und vor allem eine optimale adaptive Sampling-Strategie. Die Besonderheit hierbei ist, dass durch die simultane Formulierung das gesamte System an Sensitivitäten höherer Ordnung weggelassen werden kann, da es in einer NLP-Formulierung möglich ist, nach allen Optimierungsvariablen direkt abzuleiten. Weiterhin wurde analytisch hergeleitet, dass die Struktur der Jacobi-Matrizen von den Nebenbedingungen besonders dünnbesetzt ist und dass die meisten Terme in den Ableitungen erster Ordnung für die Berechnung der Ableitungen höherer Ordnungen wiederverwendet werden können. Zusammengefasst wurde ein allgemeingültiger Ansatz formuliert, mit dem im Gegensatz zum konventionellen sequentiellen Ansatz höhere optimale Versuchsplanungsanwendungen gelöst werden können mit flexibleren Problemformulierungen und robusteren Lösungen.

Die Anwendung auf einen theoretischen instabilen CSTR Prozess mit Kopplung der Energiebilanz hat gezeigt, dass mit dem vorgestellten Ansatz ein optimales Ergebnis,

während mit kommerziellen Software-Tools, die den Stand der Technik verkörpern, keine zulässige Lösung gefunden werden konnte. Mit einer weiteren Anwendung auf ein Schätzproblem von komplexen Reaktionskinetikparametern einer Rhodium-katalysierten Hydroformylierung wird veranschaulicht, dass im Gegensatz zu konventionellen Versuchsplanungsmethoden der experimentelle Aufwand drastisch reduziert werden konnte bei besseren Parameterwerten und Konfidenzintervallen.

ABSTRACT

Model-based optimal experimental design gains more and more impact through the last two decades, especially regarding applications in chemical, bio-chemical and process engineering. Experiments are optimized by planned control trajectories so that maximum information can be provided with minimum experimental effort. Hence, there is an enormous potential for saving time and resources in most application cases. The simultaneous optimization approach is presented which describes an efficient strategy for solving the increasing complex problem formulation to optimal experimental design. The main contribution of this thesis is the mathematical derivation of optimal experimental task formulated as a simultaneous optimization problem, whose crucial point lies in the total discretization of the dynamic system model as well as the sensitivities equations. Thus the original optimal control problem is converted to a NLP problem which can be solved more efficient and more robust via global optimization algorithms. Furthermore the proposed approach covers all requirements of an advanced optimal experimental design problem. These are in particular flexible application of control trajectories of different orders, optimization of the initial conditions, strict compliance to process constraints with respect to the controls and especially the state variables, limitation on the sampling number and experiment duration and especially an adaptive optimal sampling strategy. The crucial advantage of the simultaneous approach is that the entire system of higher order sensitivities can be left out, since all equation constraints can be directly differentiated with respect to the optimization variables. Furthermore it has been analytically derived, that the structure of the Jacobian of the constraints are extremely sparse and most parts of the first-order constraint derivatives can be reused for the calculation of derivatives of higher orders. Altogether, a universal approach has been created so as to solve more complex optimal experimental design tasks providing full flexibility of the problem formulation and robustness of the results.

The application on an instable CSTR process including the energy balance has shown that an optimal result is obtained with the simultaneous approach whereas state-of-the-arts commercial software tools cannot find any feasible solution. Moreover, the application on a parameter estimation problem of complexed reaction kinetic parameters of a Rhodium-catalyzed hydroformylation reaction process shows that in contrast to the conventional approach of design of experiments the experimental effort

could be drastically reduced at better estimated parameter values and confidence intervals.

Contents

LIST OF FIGURES

LIST OF TABLES

NOTATION

Latin letters

c	equation constraint
c_i	concentration of substance i $\left[\frac{g}{l}\right]$ or i-th diagonal element of the covariance matrix C_θ
g	model equation
h	decision interval length
i	index
k_{ref}	reference frequency factor
k	collocation index
k_q	collocation index of the q-th control
l	collocation element index
m	measurement function
n	mole amount [mol]
p	pressure [bar]
q	collocation order of the control
$\dot{r}$	reaction rate $\left[\frac{mol}{l \cdot s}\right]$
s	sensitivity
t	time
t_f	final end time
t_{α,N_m-N_θ}	student's distribution with respect to α and $N_m - N_\theta$
u	control variable
x	summarized state variables
y	measured state variable
$\hat{y}$	measurement of y
z	state variable

A_R	area of the reactor jacket
B	mass matrix
C_θ	covariance matrix
E	variance operator
E	expected value operator
E_a	activation energy $\left[\frac{J}{mol}\right]$
F	measurement matrix or flow rate
F_{in}	inlet mole flow rate
F_{out}	outlet mole flow rate
H	Hessian
H_0	reference Henry constant $\left[\frac{bar \cdot l}{mol}\right]$
H_{ads}	adsorption enthalpy $\left[\frac{J}{mol}\right]$
ΔH_r	reaction enthalpy $\left[\frac{J}{mol}\right]$
I	Identity matrix
J	Jacobian
K	mechanistic kinetic parameter or collocation order
$K_{u,q}$	collocation order of the control u with order q
L	Likelihood function or reactor level
M_F	Fisher Information matrix
N	number
N_d	number of decision intervals
N_e	number of finite element
N_{sp}	total number of possible sampling decisions
$N_{sp,i}$	sampling decision number for state i
N_u	number of control variables
N_z	number of state variables
N_θ	number of model parameters

S_{sp}	dynamic sensitivity matrix
T	reactor temperature or transpose operator [K]
T_{in}	temperature of inlet stream [K]
T_{out}	temperature of outlet stream [K]
T_{ref}	reference temperature [K]
U	heat transfer coefficient
V_L	volume liquid phase [l]
X	summarized optimization variable

Greek letters

θ	parameter
$\hat{\theta}$	estimated parameter
θ^*	"true" or "optimal" parameter
ϵ	measurement errors
σ	measurement variance
κ	diagonal element of covariance matrix
ω	sampling decision
τ	element local time
φ	Lagrange polynomials
λ	Lagrange multiplier
δ	model divergence
Φ	objective function
Σ	measurement covariance matrix

Indices

d	decision
i	state

j	parameter
k	collocation points
l	element
d	decision
m	measurements
q	control
B	biomass
Doce	1-dodecene
Doca	dodecane
E	equality constraints
HydA	hydrogenation reaction
HyfoA	hydroformylation reaction
I	inequality constraints
IsoA	Isomerization reaction
IsoB	Isomerization back reaction
IsoDoce	iso-dodecenes
S	substrate
α	confidence measure
θ	with respect to the parameters

Abbreviation

CSTR	Continuous Stirred Tank Reactor
DAE	Differential Algebraic Equation
FDM	Finite Difference Method
FEM	Finite Element Method
FIM	Fisher Information Matrix

GC	Gas Chromatography
IP	Interior Point
LSE	Least Square Estimation
MLE	Maximum Likelihood Estimation
NLP	Nonlinear Programming
OASE	Optimal Adaptive Sampling Strategy
OCFEM	Orthogonal Collocation on Finite Elements Method
OED	Optimal Experimental Design
OP	Optimization Problem
PDE	Partial Differential Equation
SQP	Successive Quadratic Programming
TDC	Tridecanal
TMS	Thermomorphic Solvent
w.r.t.	with respect to

Special characters

$*$	optimal condition
$\otimes$	Hadamard product
vec	vectorization operator
$\lvert M \rvert$	determinant of matrix M
$\underline{v}$	vector v
$\underline{\underline{M}}$	Matrix M
∇	Nabla operator
∇_{xy}	Laplace operator with respect to x and y

1 Introduction

1.1 Problem and objectives

In the field of experimental design, the name Fisher has more impact than anyone else since the 1920s. Terminologies like Fisher distribution, maximum likelihood and especially Fisher information matrix (FIM) are still characterizing the work of the experimental design community. It is Fisher's insight that is the driving force of this work that instead of accepting the fate that experiments which were carried out has no impact for a posteriori data analysis, one should a priori include the methods of design of experiments so as to avoid meaning less experiment results. This leads to considering parameter uncertainties and information content of experiments at the very early stage of a process investigation or more generally at every process development step.

Since the last two decades concepts of design of experiments have been carried from linear dependencies to dynamic nonlinear models especially in the field of chemical engineering. Classical methodologies of statistical design of experiments such as factorial design, response surface method or more generally linear regression methods reach their limits when it comes to complex reaction kinetics, thermodynamic and mass transfer phenomena. Here engineers mostly are faced with complex processes which have nonlinear and time-dependent characters. In contrast to the parameters of linear regression models which are in general unbounded scalars, their model parameters have physical meanings and therefore obey physical limitations. On the other hand, there are more degrees of freedom because of their time-dependent character. The control or perturbation decisions are not only at the beginning of the process and/or constant but can rather be dynamically changed throughout the entire process. This leads to a formulation of dynamic design of experiments as an optimal control problem which is often referred in literature as optimal experimental design (OED). The degrees of freedom are time-dependent control trajectories and the objective functions are function of the dynamic FIM in case of improving posteriori parameter estimations, or function of the divergence of competing models in case of model discrimination. This work focusses on improving parameter estimations but the methodology can be extended to the latter case in a straightforward way.

The step to time-dependent processes brings not only more degrees of freedom but also new challenges. First, the choice of the optimal initial conditions is crucial for the design of batch and semi-batch experiments, in particular regarding the estimation of reaction kinetic parameters. Secondly, the order of the dynamic control trajectories should be flexible. In chemical engineering control types of zero-order (step-control) and first-order (ramp-control) have been found as most relevant since they cover most of the control structures of chemical processes, for example flow rate, temperature, pressure, stirrer speed, inlet concentration etc... Control types of higher orders are generally unwanted since tracking of their trajectories is hardly to realize in practice. Another important issue is related to constraints on control and state variables. The first issue mostly reflects technical limitations of the control structure of the process. Whereas the second point is crucial for product specifications and especially for the process safety which basically is the most important issue regarding process operation. Therefore a strict compliance to the constraints on the state variables is indispensable. Furthermore, a proper sampling strategy has a big impact on the subsequent parameter estimation results since parameter estimation are carried out based on measurements which has to be taken during the experiments. Therefore, the OASE has been introduced which properly gives the answer to the questions of “where” and “when” to measure. Finally, limitations on the sampling numbers of each measured variable and the total experiment time are also essential.

Based on the discussed requirements above an advanced OED formulation has been defined for this work which is illustrated in Figure 1.1:

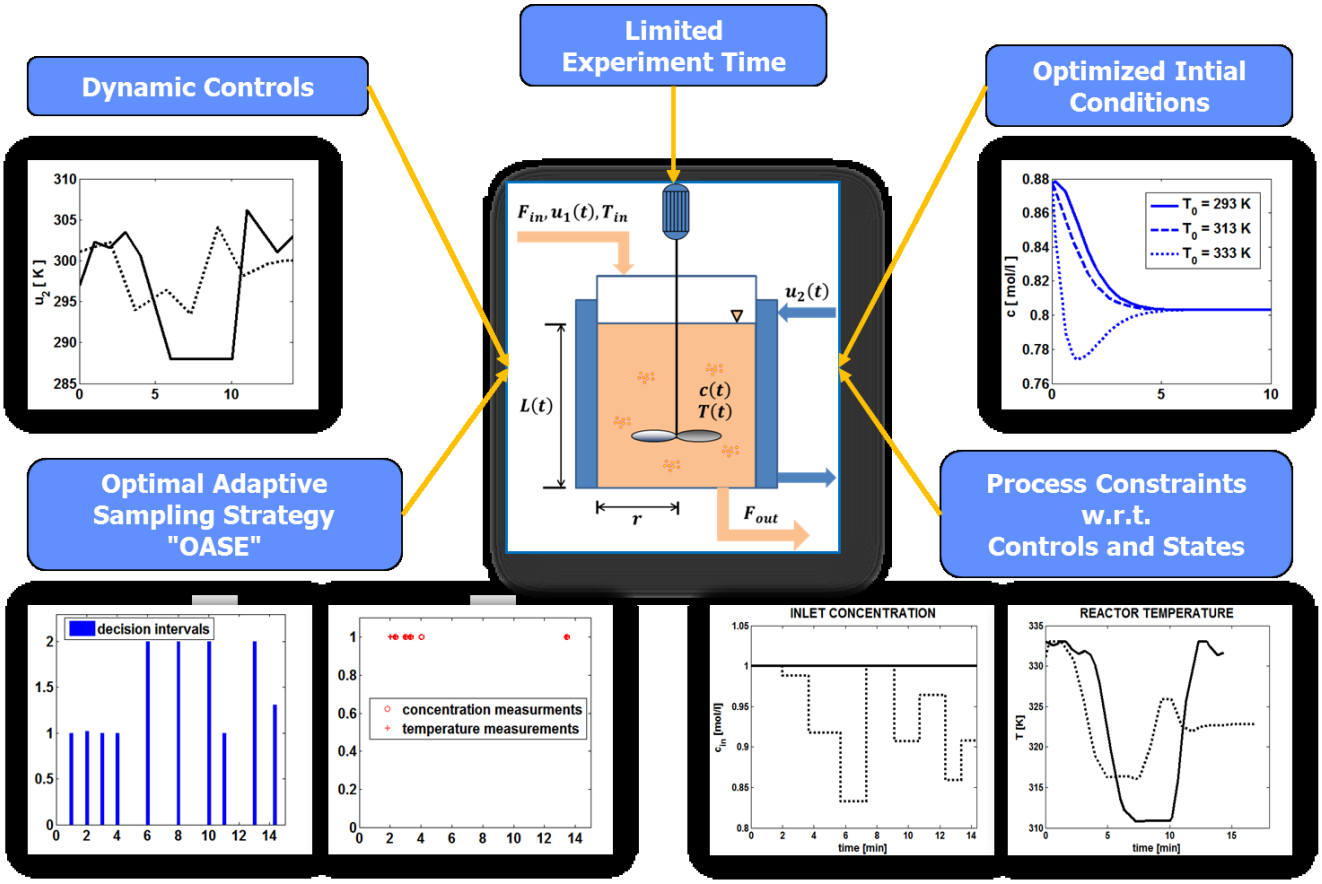


Figure 1.1 Advanced OED - formulation

The summarized requirements are:

- Optimization of the initial conditions
- Dynamic control trajectories of different types (steps, ramps, ...)
- Strict compliance to process constraints (states and controls)
- Optimal adaptive sampling strategy (OASE)
- Limitation on sampling number and experiment duration

The simultaneous optimization approach to OED represents the key contribution in this thesis, which has the advantage over the conventional approaches that it covers all requirements of an advanced OED formulation (Figure 1.1). A more detailed introduction to the differences between conventional and the new approach is given in the next subsection following by an engineering application, which aims to motivate

and lead the reader to the field of OED and the related optimization strategies. After the introduction in chapter one, chapter two presents the fundamental theories on which OED is built on while giving a short introduction to the theory of statistical analysis regarding the parameter estimation problem. Furthermore, the solution of the given OED problem is formulated via the simultaneous optimization approach and the derivation of the derivatives of the objective function as well as the constraints. Chapter three gives a detailed insight to the solution and implementation strategies of the presented optimization approach. Strength and advantages of the applied solution strategy is described step by step. Two application examples are given in chapter four. The first application example is related to an unstable CSTR-process whereas the second application example is related to a complex estimation problem of reaction kinetic parameters. The last chapter summarizes the thesis and gives outlooks and suggestions for future investigations.

1.2 Relevance of the work

There are only few developed program packages which solve with different strategies the presented OED task. State-of-the-art programs are dynamic optimization methods with embedded DAE solvers which are referred here as sequential approaches. Backbone of these methods are efficient DAE integration algorithms, which solve the system with the associated sensitivities-state equation system as initial value problems (Barz et al. 2011; Li & Petzold 1999; Albersmeyer & Bock 2009). The subsequent connection of the calculated state variables and gradient information to NLP-optimizers can be carried out in a straightforward way. A distinction is made here between the single shooting and the multiple shooting approaches.

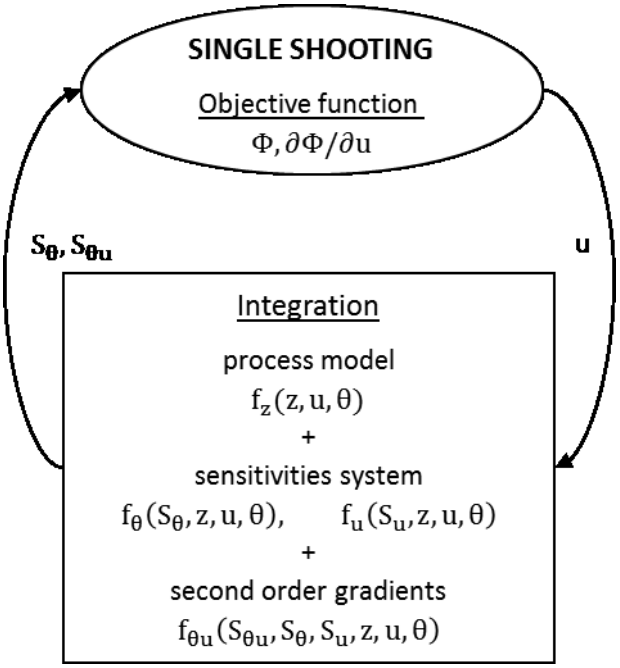


Figure 1.2 Single shooting optimization approach

It is well known that single shooting methods (Figure 1.2) generally has troubles with unstable systems, since computed control outputs during the optimization procedure may lead the connected DAE-solver into unstable regions, where convergence is not achieved because of unbounded trajectories. Out of this reason, multiple shooting methods were introduced in the past (Bock & Pitt 1984; Bauer et al. 2000). Basically, the functions of the state variables are divided into small time periods, in which the models are then also solved by DAE-solvers (Figure 1.3). In contrast to the single shooting approach, instability and poor conditioning of the problem can then be overcome since additional inequality constraints can be formulated so as to set bounds on state variables. However, this feature only yields for the endpoints of each element. All sequential approaches are based on DAE solvers, and thus, the solution of the extended sensitivities-state equation system has to be integrated in each optimization step. This can require an extensive computational load since the integration of state variables and target sensitivities constitute the main costs of the optimization process, in particular if the system model includes many states, parameters, and especially control variables. Another common disadvantage of both methods relates to constraints on state variables, which therein can only be considered indirectly by approximations. Further disadvantages are shown in (Cervantes & Biegler 2000; Shivakumar & Biegler 2006) in

more detail while discussing path constraints and dense block structure with respect to constraint gradients.

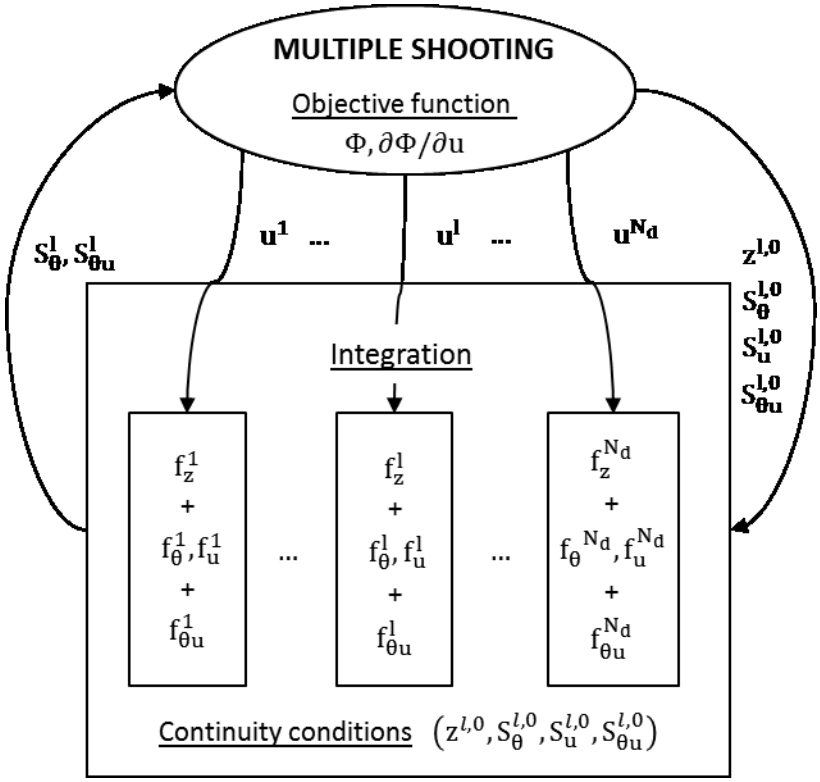


Figure 1.3 Multiple shooting optimization approach

On the other hand there are "direct transcription" methods which lead to a simultaneous approach (Figure 1.4) (Biegler 2010). The crucial point of the simultaneous formulation is the total discretization of the sensitivities-state equation system into equality constraints resulting in a NLP-optimization problem. Hereby it is possible to calculate the derivatives of the experimental design criteria directly as functions of the discretized sensitivity variables and to exploit the sparse-structure of the constraint derivatives to a full capacity. Inequality conditions for state and control variables can then directly be embedded into the formulation over the whole domain and therefore can overcome convergence difficulties of sequential optimization approaches. A further advantage is the implementation of control functions of flexible order, which can be treated straightforwardly due to the same discretization scheme as the state variables.

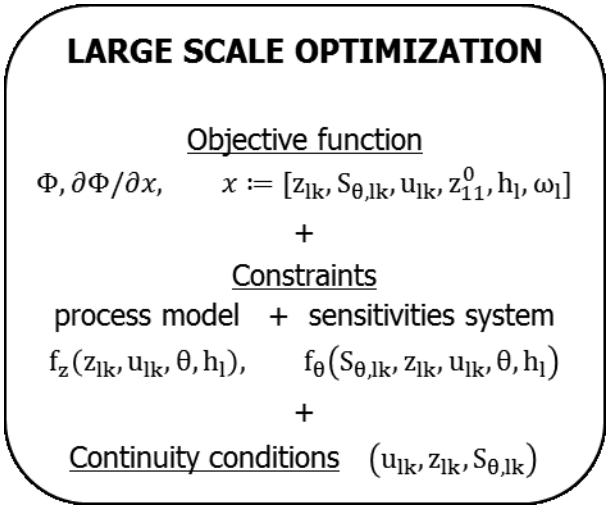


Figure 1.4 Simultaneous optimization approach

It is worthwhile to point out that there are two state-of-the-art software tools, which solve OED tasks. The first one is the experimental design package of the commercial software gPROMS, which uses the sequential optimization approach (Galvanin et al. 2007; Franceschini & Macchietto 2007; Galvanin et al. 2009). As pointed out, the main drawback of this approach is that it cannot handle unstable systems properly (see section 4.1). Furthermore it is possible to choose in gPROMS controls of different types (zero- and first-order), but it is not possible to include the continuity condition for first-order controls. Hence the problems coming with the implementation of discontinuous control trajectories cannot be overcome with this package. The second one is VPLAN, which is developed by the "BASF Junior Research Group Optimum Experimental Design" of the University Heidelberg using the multiple shooting optimization approach (Schöneberger et al. 2009; Körkel et al. 2004; Arellano-Garcia et al. 2007). Up to now, it only supports optimization results with zero-order controls.

1.3 An introduction application example

Goal of process modeling tasks is to have a trustful mathematical description which means a feasible model with a high confident parameter set (Englezos & Kalogerakis 2000). In the optimal case, the model should describe the process within the entire operating space rather than only individual operating points. In other words, assuming that an a priori known model structure is the best description of the process mechanisms, its parameters must be valid and trustful for every possible operating condition. This aim can be accomplished by the methods of experimental design so as to

maximize the parameter accuracy or in other words minimizing their uncertainties (Fisher 1971; Steinberg & Hunter 1984; Montgomery 2001). Without loss of generality a comparison between the factorial design, the sequential and the simultaneous optimization approach to OED is discussed based on the next illustrating example so as to introduce the advantages of the developed simultaneous optimization approach.

For this purpose, the reactor of a biomass reaction process is shown in Figure1.5. The considered model consists of two differential equations and a reaction rate, which together describe a biomass population and its consumption of substrate (Espie & Macchietto 1989) (as seen in equations (1.1) to (1.3)). The system state variables are the biomass concentration c_B and the substrate concentration c_S. The reactor is fed by a substrate stream which provides two control possibilities. The first one is addressed as the dilution factor u_1, which can be regarded as the feed stream and the second one is the inlet substrate concentrationu_2. It is assumed that there are two unknown model parameters θ_1 andθ_2. Both have an initial guess of 0.5. The parameters θ_3 and θ_4 are fixed in this case study to their reference values 0.05 and 0.55 respectively.

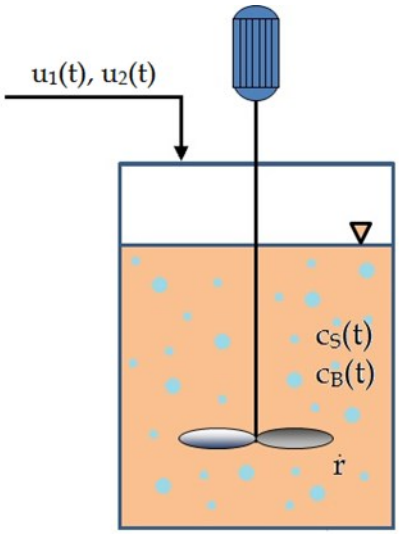


Figure 1.5 Semibatch biomass reactor

$$\frac{dc_B}{dt} = \dot{r} - (u_1 + \theta_4)c_B \qquad (1.1)$$

$$\frac{dc_S}{dt} = -\frac{\dot{r}}{\theta_3} + (u_2 - c_S)u_1 \qquad (1.2)$$

$$\dot{r} = \theta_1 \frac{c_B c_S}{\theta_2 + c_S} \qquad (1.3)$$

Technical limitations of the system have to be adhered to the following constraints on state and control variables:

$$1\ \mathrm{g/l} \leq c_B \leq 25\ \mathrm{g/l}, \qquad 0.01\ \mathrm{g/l} \leq c_S \leq 25\ \mathrm{g/l} \tag{1.4}$$

$$0.05\ \mathrm{h^{-1}} \leq u_1 \leq 5\ \mathrm{h^{-1}}, \qquad 0.2\ \mathrm{g/l} \leq u_2 \leq 35\ \mathrm{g/l} \tag{1.5}$$

$$1\ \mathrm{g/l} \leq c_{B,0} \leq 25\ \mathrm{g/l}, \qquad 0.1\ \mathrm{g/l} \leq c_{S,0} \leq 25\ \mathrm{g/l} \tag{1.6}$$

The lower bounds on the state variables guarantee a minimum population of the biomass and a minimum substrate amount so as to feed the biomass. The bounds on the control variables are taken from (Asprey & Macchietto 2002). The aim of the experimental design task is to minimize the parameter uncertainties (see section 2.2) by running appropriate experiments, such that a best possible identification of the parameters $\theta_{1,2}$ can be achieved afterwards.

1.3.1 Factorial design

A well-known method is the factorial design approach which has its roots from the statistical experimental design analysis (Montgomery 2001).The crucial point of the methodology is to divide the space of the manipulating variables into constant levels independently of whether they are continuous or discrete. The most common strategies are the 2k-design, in which the manipulating variable space is mapped to a LOW and a HIGH level and the 3k-design, where an additional MEDIUM level is included. For example, applying a 2^k-design to the continuous controlu_1, one obtains u_1^L = 0.05 h^{-1} andu_1^H = 5 h^{-1}. A 2k-design is applied to the process controls u_1 and u_2 leading to 2^2 combinations and the 3^k-design is applied to the initial conditions $c_{B,0}$ and $c_{S,0}$, and thus, leading to 3^2 combinations as seen in Table 1. This results in an experiment design with 36 experiments.

Table 1 Design conditions – biomass reactor

	LOW	MEDIUM	HIGH
u_1	$0.05\ h^{-1}$	–	$5\ h^{-1}$
u_2	$0.2\ g/l$	–	$35\ g/l$
$c_{B,0}$	$1\ g/l$	$12\ g/l$	$25\ g/l$
$c_{S,0}$	$0.1\ g/l$	$12\ g/l$	$25\ g/l$

Since the factorial design method only takes into account the manipulating variable space but neither the model structure nor the process constraints, 27 (75%) of the planned experiments violate the constraints on the states variables as listed in Table 31 (appendix). The 9 remaining admissible experiments are carried out for the subsequent parameter estimation task in the "real reactor" which is represented by the process model with the true parameter set $[\theta_1 = 0.31; \theta_2 = 0.18]$. The admissible experiment "L-L-L-H" with the according conditions is presented as an example in Table 2. Since the sampling policy is not taken into account in the factorial design approach the samplings are chosen to be equidistant.

Table 2 Factorial design – admissible experiment "L-L-L-H"

process controls		process variables	
$u_1(t) = \text{const.}$	$0.05\ h^{-1}$	$c_B(t = 0)$	$1\ g/l$
$u_2(t) = \text{const.}$	$0.2\ g/l$	$c_S(t = 0)$	$25\ g/l$
experiment duration	$10\ h$		
sampling vector [h]	$[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$		

The subsequent parameter estimation results were obtained from the gPROMS's parameter estimation toolbox as shown in Table 3.

.

Table 3 Parameter estimation results – Factorial design

	Initial guess	Final value	Standard deviation
θ_1	0.5	0.31001	0.0003114
θ_2	0.5	0.18024	0.006377

The standard deviations reveal properly estimated parameters which are very close to the real ones $\theta_1^* = 0.31$ and $\theta_2^* = 0.18$. Nevertheless, the price to achieve that was too high since 75% of the experiments were useless and 9 experiments with each 10 hours experiment time is still far from optimal experimental conditions. Therefore, one single optimized experiment is used in the next section so as to emphasize the difference between factorial design and OED.

1.3.2 Optimal experimental design

The crucial advantage of the OED approach is that the processes are dynamically addressed, and thus, exploiting their time-dependent character. Therefore, optimal control trajectories are used instead of constant control levels. Furthermore, the formulation of the OASE, which gives the answer to "where" and "when" to measure, can be applied in a straightforward way. Applying this concept to the factorial design approach would give thousands experimental setups, which is simply impossible to carry out in an economically meaningful way, thus, the problem is rather formulated as an OED problem.

Sequential optimization approach

First, the OED task was solved by the commercial software gPROMS as the reference case, which uses the sequential optimization approach (see 2.4.2). The dilution factor u_1 is chosen to be a linear function (first order), whereas the inlet concentration u_2 is chosen to be a step function (zero-order). Furthermore, there are 10 time varying decision intervals in which the control profiles are optimized. The process constraints (1.4) - (1.6) are included in the optimization task.

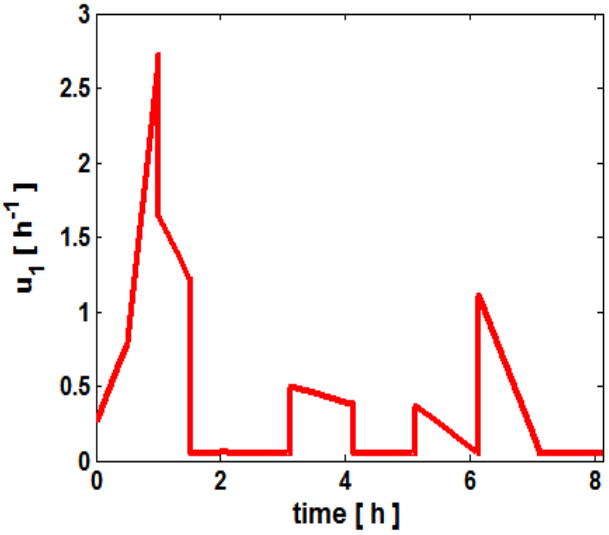


Figure 1.6 Control trajectory of the dilution factor, sequential approach

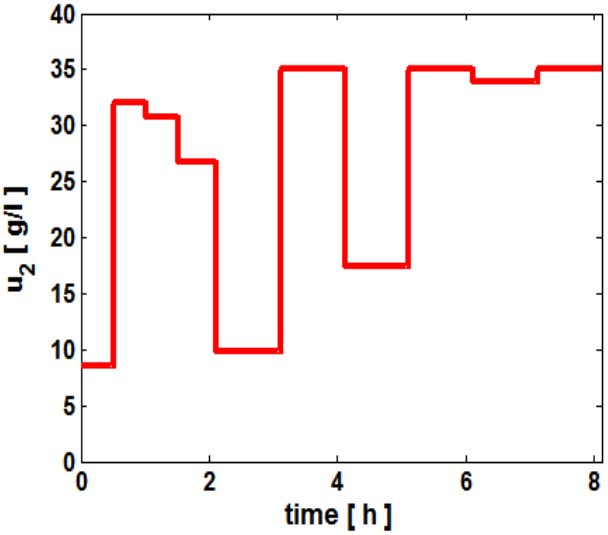


Figure 1.7 Control trajectory of the inlet concentration, sequential approach

The optimized control trajectories for u_1 and u_2 can be seen in Figure 1.6 and Figure 1.7. The discontinuities in the control profile of u_1 are especially noticeable. Although the commercial tool gPROMS allows choosing between zero and first-order controls, there is no way to adjust the continuity condition for the control trajectories, and thus, the result for u_1 can hardly be implemented in practice. Furthermore, the resulting profile for $u_2$is not optimal. Sinceu_2 linearly enters the system formulation, it has to hit its bounds (bang-bang solution) for optimality (see 3.5). The corresponding state

variables profiles are presented in Figure 1.8 and Figure 1.9 including the measurements resulting from the OASE.

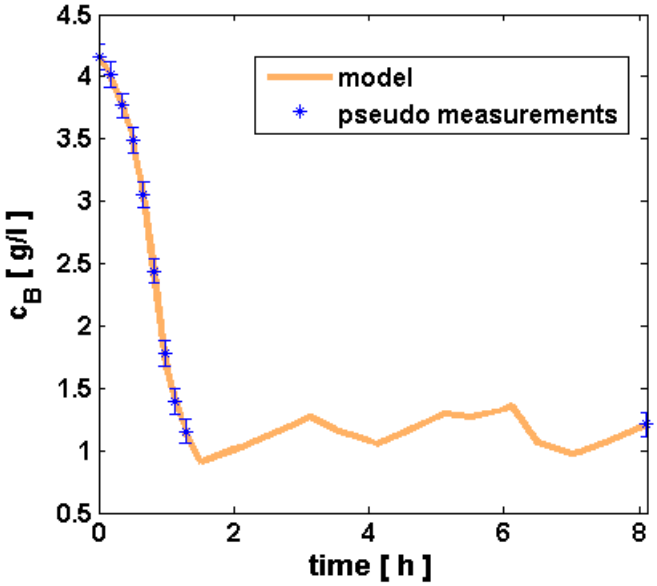


Figure 1.8 Concentration profile of the biomass, sequential approach

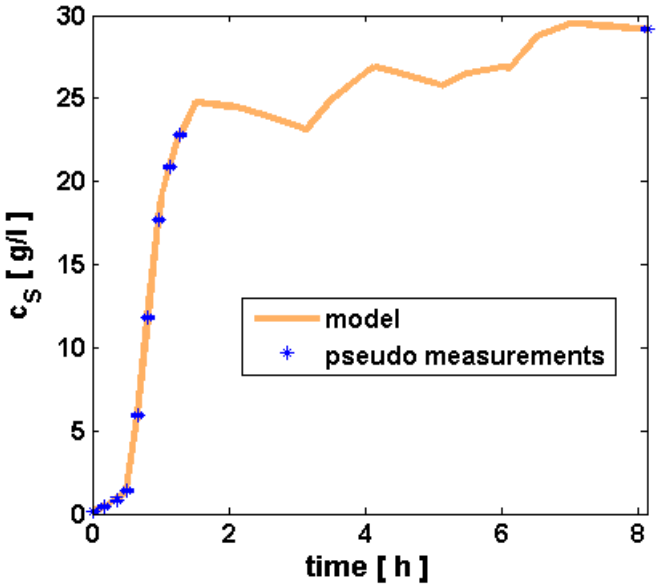


Figure 1.9 Concentration profile of the substrate, sequential approach

The profile for the substrate concentration overshoots its upper bound and thus clearly violating the process constraints. This is another drawback resulting from the sequential optimization approach since constraints on the state variables can only be considered by approximations. Moreover the optimized results for the initial conditions of c_B and c_S are also not optimal since they also linearly enters the system formulation but do not hit

their bounds (see 3.5). The subsequent parameter estimation for this case is shown in Table 4.

Table 4 Parameter estimation results – Sequential approach

	Initial guess	Final value	Standard deviation
θ_1	0.5	0.30993	0.005405
θ_2	0.5	0.17963	0.05735

The estimated parameter values and their standard deviations are slightly worse than the results obtained from the factorial design. However, it has to be considered that only one experiment was needed for matching the real parameter values with acceptable deviations.

Simultaneous optimization approach

Although state-of-the-art commercial tools with design of experiments packages basically address the OED problems, they still lack in terms of crucial optimization aspects, for example parameterization of the controls, process constraints and "global" optimality since they are formulated based on sequential optimization approaches. The results for the optimal control trajectories via the simultaneous optimization approach as shown in Figure 1.10 and Figure 1.11 have been obtained from the implementation with the global optimizer BARON in GAMS (see section 3.3 and 3.4 in more detail) (Neumaier et al. 2005). The implementation of different control order can be carried out in a straightforward way including continuity conditions due to the simultaneous formulation, see OP2 in equation (2.58).

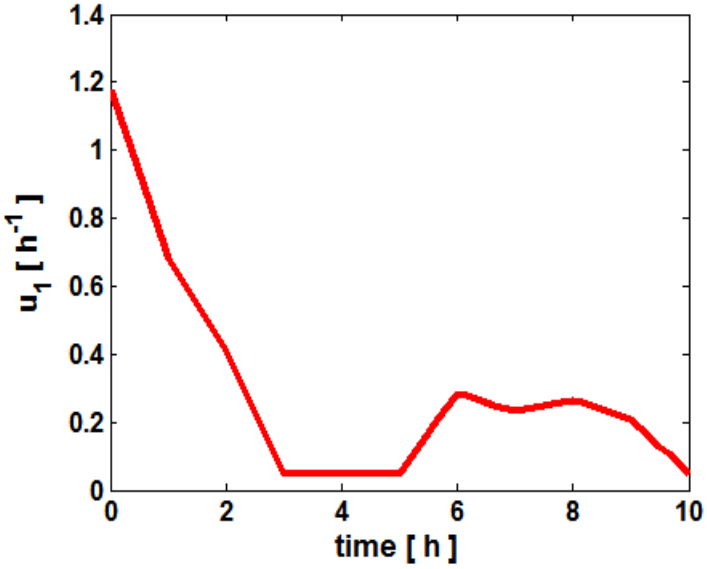


Figure 1.10 Control trajectories of the dilution factor, simultaneous approach

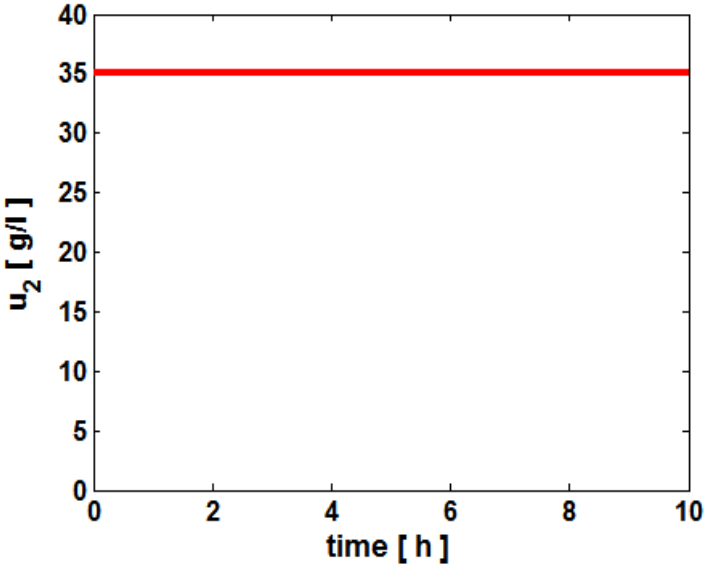


Figure 1.11 Control trajectories of the inlet concentration, simultaneous approach

Moreover, the optimal trajectory regarding u_2 is optimal since it results in a bang-bang solution. Even if the u_2 is set in this case to a linear type, the outcome still hits the upper constraint bound as shown in Figure 1.11.

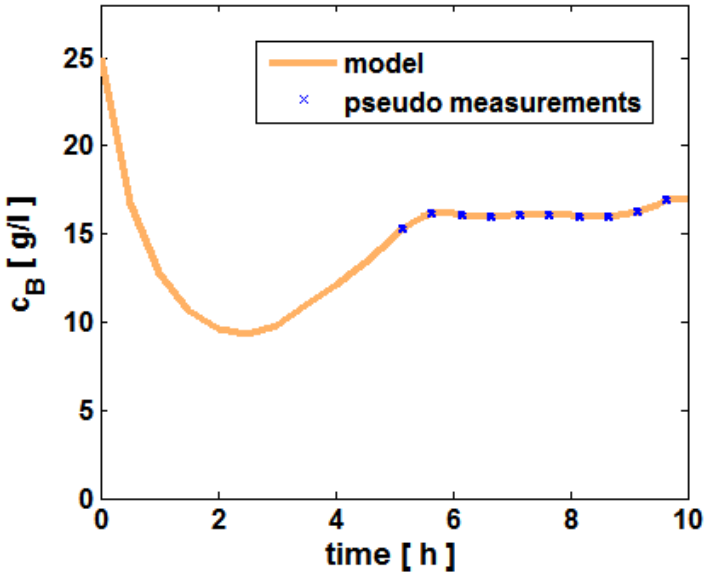


Figure 1.12 Concentration profile of the biomass, simultaneous approach

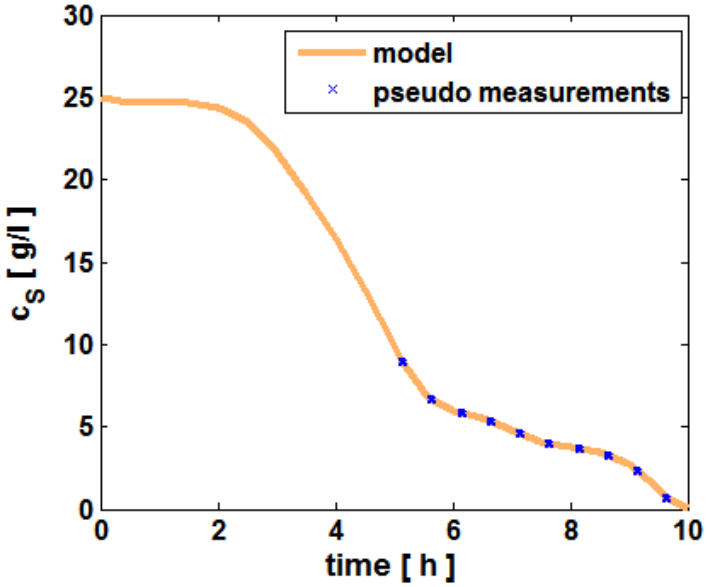


Figure 1.13 Concentration profile of the substrate, simultaneous approach

The corresponding profiles of the state variables are shown in Figure 1.12 and Figure 1.13 including the measurements resulting from the optimal sampling strategy. The concentration profiles fulfill the process constraints in both cases. In contrast to the sequential approach, the optimized results for the initial conditions of c_B and c_S hit their upper bounds with $c_B(t = 0) = 25\ g/l$ and $c_S(t = 0) = 25\ g/l$, and thus, fulfilling optimality. The subsequent parameter estimation for this case is shown in Table 5.

Table 5 Parameter estimation results – Simultaneous approach

	Initial Guess	Final value	Standard deviation
θ_1	0.5	0.31000	0.0004851
θ_2	0.5	0.18060	0.01578

The comparison between the three investigated methods is given in Table 6.

Table 6 Parameter estimation results – comparison between all approaches

	Factorial design	Sequential approach	Simultaneous approach	Reference value
θ_1 Final value	0.31001	0.30993	0.31000	0.31
θ_1 Standard deviation	0.00031	0.00541	0.00049	-
θ_2 Final value	0.18024	0.17963	0.18060	0.18
θ_2 Standard deviation	0.00638	0.05735	0.01578	-

From this small scale introductory example it becomes clear that if it comes down to processes, which has a dynamic character and where process constraints has to be strictly obeyed, the OED formulation is superior to the classical factorial design approach because of two main reasons. The first point is that OED approaches address the problem dynamically, and thus, gaining more information from the dynamic process behavior (Galvanin et al. 2011; Telen et al. 2012). The second point is the lack of the factorial design to include process constraints into the problem formulation. Furthermore, this example also reveals that the simultaneous optimization formulation reaches better results compared to the sequential approach. One of the important aspects is that global optimizers can handle the simultaneous optimization formulation much better than a sequential formulation, as seen by the presented results above. Another important point is that strict constraints on state variables of the process can be handled more properly as demonstrated by the application study in section 4.1.

2 Experimental design – state-of-the-art and theoretical background

The fundamental problem of the investigated OED approach is a parameter estimation problem. Let $\underline{\hat{y}}$ be the vector of measurement values from the process state variables of interest, $\underline{y}$ its corresponding model description and $\underline{\delta} := \underline{\hat{y}} - \underline{y}$ the difference or the distance between $\underline{\hat{y}}$ and $\underline{y}$. Then, the most common formulation for the parameter estimation problem represents the minimization of the scalar product $\underline{\delta}^T\underline{\delta}$, which is also known as the method of least squares[1] (LSE) (Geer 2008; Raol et al. 2004).

$$\underline{\delta}^T\underline{\delta} = \left(\underline{\hat{y}} - \underline{y}\right)^T \left(\underline{\hat{y}} - \underline{y}\right) \rightarrow \min \qquad (2.1)$$

In chemical engineering, $\underline{y}(t, \underline{\theta}, \underline{u}, \underline{z})$ generally is modeled as a function of time t, the parameters θ, the manipulated variables u, and sometimes a linear combination or nonlinear function of other state variables $\underline{z}$. Thus, it is plausible that, how the experiments are controlled, will fundamentally affect the outcome of the subsequent parameter estimation procedure. Furthermore, statistical analysis with respect to confidence regions of the estimated parameters is crucial for every proper estimation results. Since the topic is very complex, this chapter aims at giving a short overview over the theory of the whole domain forging a bridge between basic terms of statistical analysis regarding parameter estimation and the formulation of the advanced OED problem. The theory of linear regression analysis constitutes the first part including basic concepts of statistical analysis. The second part draws a line to the field of nonlinear regression analysis and finally the third part presents the derivation of the OED formulation.

2.1 Linear regression

In the linear regression analysis the model functions are linear regarding the parameters $\underline{\theta}$ and also called the response surface (Box & Draper 1987; Myers et al. 2009; Khuri & Mukhopadhyay 2010).

[1] Other objective functions of the minimization task are for example the Euclidean norm $\left\|\underline{\hat{y}} - \underline{y}\right\| \rightarrow min$ or the maximum norm $\left\|\underline{\hat{y}} - \underline{y}\right\|_\infty \rightarrow min$

$$\begin{aligned} y\left(x_{1,\dots,N_x}, \theta_{1,\dots,N_\theta}\right) &= \theta_1 + \theta_2 f_1\left(x_{1,\dots,N_x}\right) + \cdots + \theta_{N_\theta} f_{N_\theta-1}\left(x_{1,\dots,N_x}\right) \\ &= \underline{\theta}^T \underline{f} \end{aligned} \tag{2.2}$$

with

$$\underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{N_\theta} \end{pmatrix}, \qquad \underline{f} = \begin{pmatrix} 1 \\ f_1\left(x_{1,\dots,N_x}\right) \\ \vdots \\ f_{N_\theta-1}\left(x_{1,\dots,N_x}\right) \end{pmatrix}$$

They can be nonlinear regarding the measurement terms $f_i\left(x_{1,\dots,N_x}\right)$ for example:

$$y\left(x_1, x_2, \theta_{1,\dots,5}\right) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 + \theta_5 x_1^2 \tag{2.3}$$

Assume that the model structure is exactly matched and $\underline{\theta}^*$ is the true parameter set then the measurement data can be represented as:

$$\begin{aligned} \hat{y}_i &= y^*\left(x_{1,i}, \dots, x_{N_x,i}, \theta^*_{1,\dots,N_\theta}\right) + \epsilon_i \\ &= \theta_1^* + f_{1,i}\theta_2^* + \cdots + f_{N_\theta-1,i}\theta^*_{N_\theta} + \epsilon_i, \qquad i = 1, \dots, N_m \end{aligned} \tag{2.4}$$

Or more compact with a matrix notation:

$$\underline{\hat{y}} = \underline{y}^* + \underline{\epsilon} = F\underline{\theta}^* + \underline{\epsilon} \tag{2.5}$$

where

$$F := \begin{bmatrix} 1 & f_{1,1} & \cdots & f_{N_\theta-1,1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_{1,N_y} & \cdots & f_{N_\theta-1,N_m} \end{bmatrix} \tag{2.6}$$

Here ϵ_i represents the measurement errors which are non-correlated zero-mean-value and normal distributed. The objective function is defined as

$$\Phi_{LSE} = \underline{\delta}^T \underline{\delta} = \left(\underline{\hat{y}} - \left(F\underline{\theta} + \underline{\epsilon}\right)\right)^T \left(\underline{\hat{y}} - \left(F\underline{\theta} + \underline{\epsilon}\right)\right) \to \min \tag{2.7}$$

The necessary condition for the minimization task is

$$\begin{aligned} \frac{\partial \Phi}{\partial \underline{\theta}} &\stackrel{!}{=} 0 \\ -2F^T\underline{\hat{y}} + 2F^T F\underline{\theta} &= 0 \end{aligned} \tag{2.8}$$

Therefore the estimated parameter vector results in

$$\underline{\hat{\theta}} = \left(F^T F\right)^{-1} F^T \underline{\hat{y}} \tag{2.9}$$

Important statistical properties of the estimated parameters are discussed in the following.

2.1.1 Expected value of the estimation

The expectation operator is applied on the estimation results in (2.9) to yield the expected value of the estimation.

$$\begin{aligned} \mathrm{E}\left(\underline{\hat{\theta}}\right) &= \mathrm{E}\left((\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\underline{\hat{y}}\right) = (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\mathrm{E}\left(\underline{\hat{y}}\right) \\ &= (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\mathrm{E}\left(\underline{y}^{*} + \underline{\epsilon}\right) = (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\left(\underline{y}^{*} + \underbrace{\mathrm{E}\left(\underline{\epsilon}\right)}_{0}\right) \end{aligned} \qquad (2.10)$$

Since the measurement errors have a zero mean value, the expected value of the estimation is the true parameter set. As a result one gets for the linear case an unbiased estimator.

$$\mathrm{E}\left(\underline{\hat{\theta}}\right) = (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\underline{y}^{*} = \underline{\theta}^{*} \qquad (2.11)$$

2.1.2 Covariance of the estimated parameters

The covariance of the estimated parameters is one important variable regarding the evaluation of the parameter estimation. It is defined as

$$\mathrm{Cov}\left(\underline{\hat{\theta}}\right) = \mathrm{E}\left[\left(\underline{\hat{\theta}} - \mathrm{E}\left(\underline{\hat{\theta}}\right)\right)\left(\underline{\hat{\theta}} - \mathrm{E}\left(\underline{\hat{\theta}}\right)\right)^{\mathrm{T}}\right] \qquad (2.12)$$

And results with (2.11) and (2.9) in

$$\begin{aligned} \mathrm{Cov}\left(\underline{\hat{\theta}}\right) &= \mathrm{E}\left[\left(\underline{\hat{\theta}} - \underline{\theta}^{*}\right)\left(\underline{\hat{\theta}} - \underline{\theta}^{*}\right)^{\mathrm{T}}\right] \\ &= \mathrm{E}\Big[\Big((\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\underline{\hat{y}} \\ &\qquad - (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\underline{y}^{*}\Big)\Big((\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\underline{\hat{y}} - (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\underline{y}^{*}\Big)^{\mathrm{T}}\Big] \\ &= \mathrm{E}\left[\left((\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\left(\underline{\hat{y}} - \underline{y}^{*}\right)\right)\left((\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\left(\underline{\hat{y}} - \underline{y}^{*}\right)\right)^{\mathrm{T}}\right] \\ &= (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\mathrm{E}\left[\left(\underline{\hat{y}} - \underline{y}^{*}\right)\left(\underline{\hat{y}} - \underline{y}^{*}\right)^{\mathrm{T}}\right]\mathrm{F}(\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1} \end{aligned} \qquad (2.13)$$

Assume for simplicity that the variances σ of all measurements errors are the same, then with Σ as the measurement covariance matrix and σ as the simplified constant measurement variance, one obtains

$$\begin{aligned} \mathrm{E}\left(\underline{\delta\delta}^{\mathrm{T}}\right) &= \mathrm{E}\left[\left(\underline{\hat{y}} - \underline{y}^{*}\right)\left(\underline{\hat{y}} - \underline{y}^{*}\right)^{\mathrm{T}}\right] \\ &= \mathrm{E}\left(\underline{\epsilon\epsilon}^{\mathrm{T}}\right) = \Sigma = \sigma^{2}\mathrm{I} \end{aligned} \qquad (2.14)$$

By substituting in (2.13), one gets the simplified form of the estimation covariance matrix C_θ

$$\begin{aligned} \mathrm{Cov}\left(\underline{\hat{\theta}}\right) &= (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{F}^{\mathrm{T}}\mathrm{F}(\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\sigma^{2} \\ &= (\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\sigma^{2} =: \mathrm{C}_{\theta} \end{aligned} \qquad (2.15)$$

Consequently, one gets for each estimated parameter θ_i the variance

$$\mathrm{D}^{2}(\theta_{\mathrm{i}}) = \mathrm{Cov}(\theta_{\mathrm{i}}, \theta_{\mathrm{i}}) = \kappa_{\mathrm{i}} \qquad (2.16)$$

with κ_i as the diagonal elements of the estimation covariance matrix C_θ or alternatively

$$\mathrm{D}^{2}(\theta_{\mathrm{i}}) = \sigma^{2}\mathrm{c}_{\mathrm{i}} \qquad (2.17)$$

with $\kappa_i = \sigma^2 c_i$, and c_i as the diagonal elements of the matrix $(F^TF)^{-1}$. In general the variance of the measurements σ is unknown and approximated with the sample variances. It is

$$\sigma^{2} \sim \mathrm{s}^{2} := \frac{1}{\mathrm{N_m}} \sum_{\mathrm{i}=1}^{\mathrm{N_m}} (\hat{\mathrm{y}}_{\mathrm{i}} - \mathrm{y}_{\mathrm{i}})^{2} \qquad (2.18)$$

Thus, it follows that

$$\kappa_{\mathrm{i}} \sim \mathrm{s}^{2}\mathrm{c}_{\mathrm{i}} \qquad (2.19)$$

2.1.3 Distribution function of the estimated parameters

A proper parameter estimation result does not only provide the estimated parameters but also the information about their confidence. In other words, how much can one trust the estimated parameter values. For this purpose it is necessary to have a closer look to the estimated parameter from the statistical point of view.

It was Gosset who stated that an estimator based on normal distributed data is not also normal but t-distributed if the necessary variance of the data σ^2 is unknown and the samplings number is small (Gosset 1908). Generally this applies to parameter estimation tasks in chemical engineering. Therefore the t-distribution allows the calculation of the distribution of the difference between the parameters and their estimators (Senn & Richardson 1994; Sheynin 1995). It becomes the normal distribution if the number of samplings reaches infinity $N_m \to \infty$ (Hogg et al. 2012). For the probability density function of a parameter θ_i as a function of its estimator $\hat{\theta}_i$ and variance κ_i is given in (2.20) and is illustrated in Figure 2.1.

$$p(\theta_i|\hat{\theta}_i, \kappa_i) = \frac{1}{\sqrt{2\pi\kappa_i}} \exp\left(-\frac{1}{2\kappa_i}(\theta_i - \hat{\theta}_i)^2\right) \qquad (2.20)$$

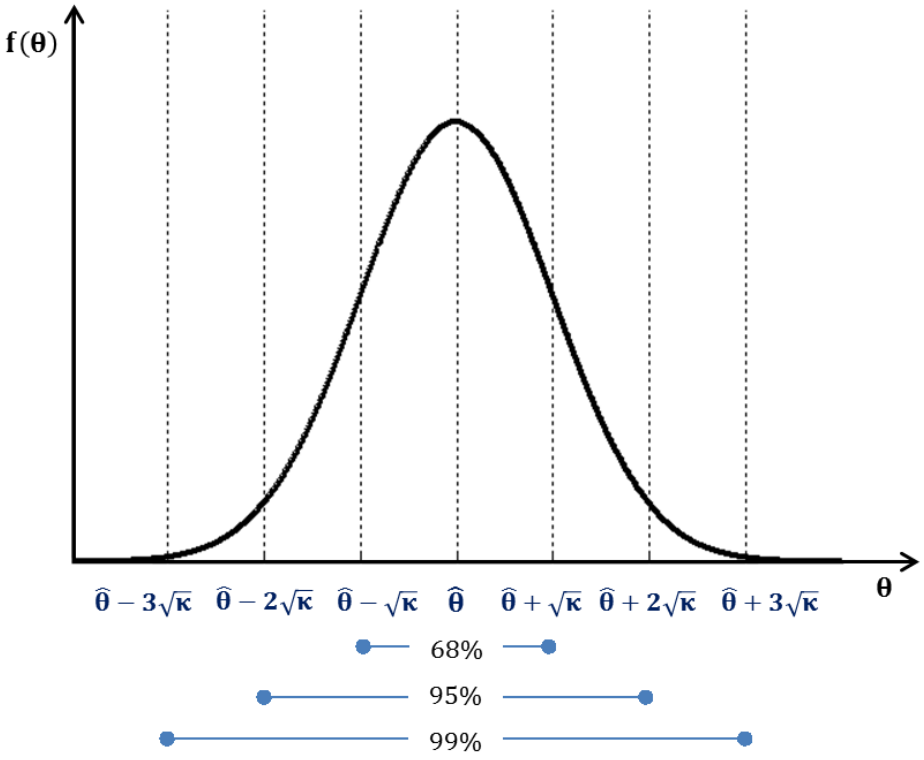


Figure 2.1 Parameter probability density function – normal distribution

It provides an intuitive knowledge about the confidence of the estimator based on the variance κ_i. The narrower the curve and the smaller the variance the more we intuitively trust the estimated parameter. However, the characterizing variance κ_i or rather σ^2 is generally unknown so that the sample variance s^2 (2.18) is used instead. Furthermore, one likes to have a statistical statement like: “The true parameter value lies with a 99%

confidence between the interval [a, b]". These constraints match the application of the t-test, which is based on the t-distribution theory as mention above. In this context, the t-values are function of the significance level α and the sampling amount N_m and determine the confidence interval of the estimated parameter values adding significance to the estimated parameters (Venables & Ripley 2003). Based on the test value

$$\frac{\left|\hat{\theta}_i - \theta_i^*\right|}{s\sqrt{c_i}} \sim t_{N_m - N_\theta} \qquad (2.21)$$

a 95% confidence interval ($\alpha = 5$) is given as

$$\hat{\theta}_i - s\sqrt{c_i}t_{\alpha,N_m-N_\theta} \leq \theta_i^* \leq \hat{\theta}_i + s\sqrt{c_i}t_{\alpha,N_m-N_\theta} \qquad (2.22)$$

The correct statistical statement based on this is: "If one continues with 100 other experiments and calculated for each the corresponding 95% confidence interval, then the true parameter value will be included in 95 of these 100 confidence intervals." Alternatively one can write for (2.22)

$$\frac{\left|\hat{\theta}_i - \theta_i^*\right|}{\sqrt{c_i}} \leq s t_{\alpha,N_m-N_\theta} \qquad (2.23)$$

Furthermore, it is more convenience to define the square of (2.23) with respect to the analysis of confidence intervals

$$\frac{\left(\hat{\theta}_i - \theta_i^*\right)^2}{c_i} \leq s^2 F_{\alpha,N_m-N_\theta} \qquad (2.24)$$

Here, $F_{\alpha,N_m-N_\theta} := t^2_{\alpha,N_m-N_\theta}$ is known as the Fisher-Snedecor- or short F-distribution (Phillips 1982; Johnson et al. 1995).

2.1.4 Confidence region

The term of confidence interval in the previous section can be generalized for the multidimensional case to a confidence region. It is

$$\left(\theta_i^* - \hat{\theta}_i\right)^T C_\theta^{-1}\left(\theta_i^* - \hat{\theta}_i\right) \leq N_\theta s^2 F_{\alpha,N_m-N_\theta} \qquad (2.25)$$

The left term in (2.25) forms a confidence ellipsoid which is illustrated below for $N_\theta = 3$.

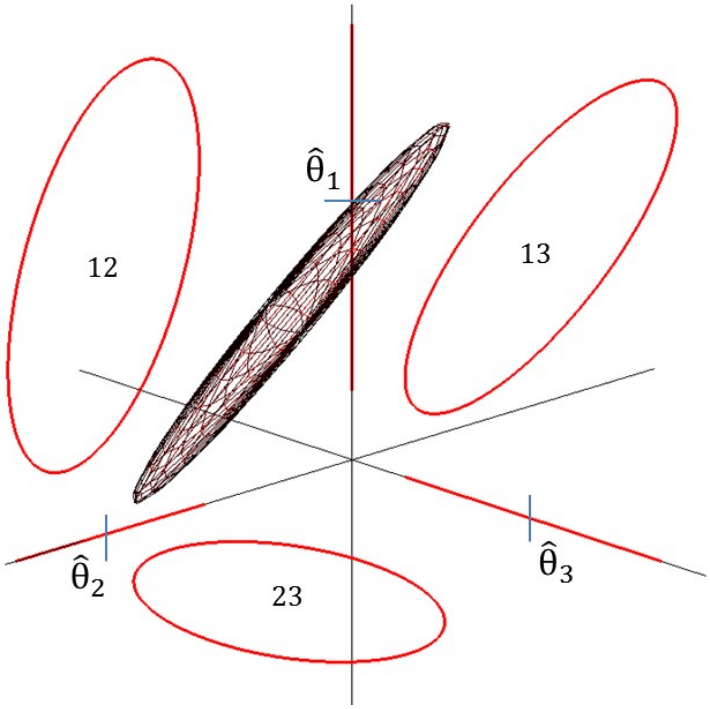


Figure 2.2 Confidence ellipsoid

As seen in Figure 2.2, there are three corresponding confidence intervals for each estimated parameter $\hat{\theta}_i$ and also three confidence region for each parameter pair. The vector of estimated parameters $\underline{\hat{\theta}}$ is a stochastic variable whose corresponding confidence region provides a concept of uncertainties. The wider the confidence region the more uncertain are the estimated parameters $\underline{\hat{\theta}}$. Therefore, the volume of the confidence ellipsoid $V(\underline{\hat{\theta}})$ is a quantitative measurement for the uncertainties of the estimation results and is denoted as follows (Bard 1973)

$$V(\underline{\hat{\theta}}) \sim \sqrt{\det C_\theta} \qquad (2.26)$$

Therefore a meaningful experimental design task is minimizing $\det C_\theta$ which is also known as the D-optimal criterion. The next section presents an overview of optimization criteria based on the concept of confidence region.

2.2 Experimental design criteria

Based on the estimation covariance matrix C_θ, the common experimental design criteria are presented in Table 7 (Steinberg & Hunter 1984; Wong 1994; Franceschini & Macchietto 2008).

Table 7 Experimental design criteria – Covariance matrix

A-criterion	$\text{trace}\, C_\theta \rightarrow \min$	minimizing the average variance (average axis radius length)
D-criterion	$\det C_\theta \rightarrow \min$	minimizing the ellipsoid volume
E-criterion	$\lambda_{max}(C_\theta) \rightarrow \min$	minimizing the largest variance (largest axis radius length)
M-criterion	$\sqrt{c_{i,max}} \rightarrow \min$	minimizing the largest side length of the enclosing box

A graphical interpretation of the presented criteria is illustrated by Figure 2.3.

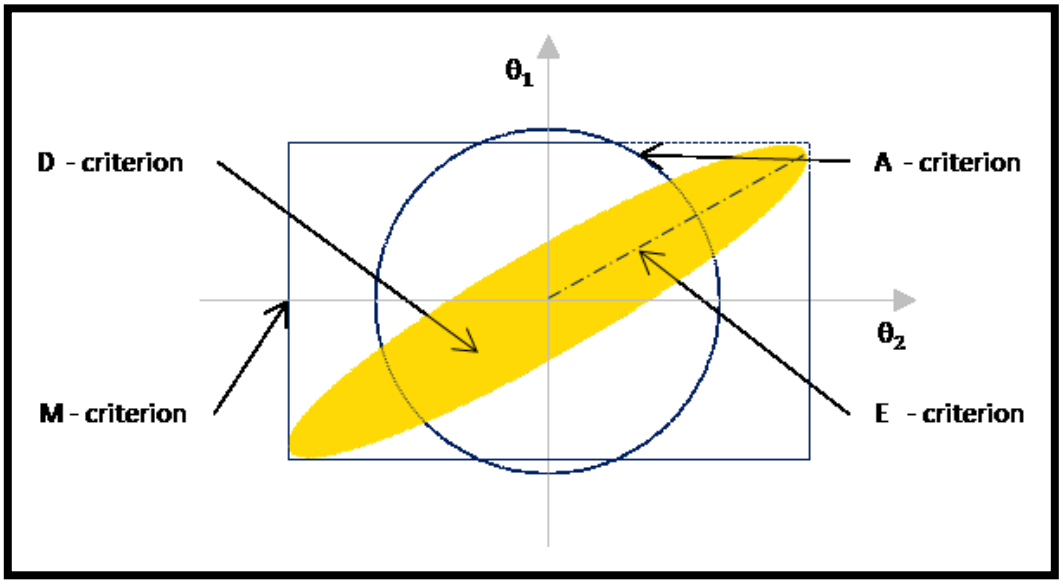


Figure 2.3 Confidence ellipsoid and common experimental design criteria for $N_\theta = 2$

The experimental design criteria can also be formulated as functions of the FIM as shown in the next sections in more detail (see equation (2.34) and subsection 2.4.3). Since the aim of this work is the introduction of the formulation of the simultaneous optimization approach to OED, the application of the methodology is focused on the most often used criteria in practice, which are the A-, D- and E-criterion (Steinberg & Hunter 1984).

2.3 Nonlinear time dependent regression

In contrast to linear regression, the process models in chemical engineering are nonlinear in the model parameters. They are generally not formulated explicitly like in (2.3) but rather represented by a DAE-system:

$$\begin{aligned} 0 &= \underline{g}\left(\dot{\underline{z}}(t), \underline{z}(t), \underline{u}(t), \underline{\theta}\right) \\ \underline{z}(t=0) &= \underline{z}_0 \end{aligned} \qquad (2.27)$$

Here $\underline{z} \in R^{N_z}$ represents the state variables, $\underline{u} \in R^{N_u}$ the control variables and $\underline{\theta} \in R^{N_\theta}$ the parameters respectively. The Maximum-Likelihood-Estimation (MLE) is generally used regarding the theory of nonlinear regression (Bard 1973; Pratt 1976; Aldrich 1997). According to the MLE, the probability that a measurement $\hat{\underline{z}}$ of data region $\delta\hat{Z}$ is represented by a certain corresponding parameter vector $\hat{\underline{\theta}}$ is given as $p\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right)\delta\hat{Z} \coloneqq L\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right)\delta\hat{Z}$ (Le 1990). A heuristic statement of the MLE is then: "The Likelihood function has its maximum at the optimal parameter $\underline{\theta}^*$, thus, among all possible parameters $\underline{\theta}$, the optimal parameter $\underline{\theta}^*$ is the most probable one which reproduces a measurement within the data region $\delta\hat{Z}$" (Ghosh & Basu 1988). Assuming that a measurement $\hat{\underline{z}}$ is normal distributed, the related Likelihood function is then defined as

$$L\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right) = (2\pi)^{-0.5}(\det\Sigma_i)^{-0.5}\exp\left(-\frac{1}{2}\underline{\delta}^T\Sigma^{-1}\underline{\delta}\right) \qquad (2.28)$$

According to equation (2.14), Σ representes the variances of the measurement errors

$$E\left(\underline{\delta}\underline{\delta}^T\right) = E\left(\left(\hat{\underline{z}} - \underline{z}\right)\left(\hat{\underline{z}} - \underline{z}\right)^T\right) \coloneqq \Sigma \qquad (2.29)$$

The Likelihood function of an experiment with N_m measurements can be given as the product of each individual measurement i.

$$L\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right) = (2\pi)^{-\frac{N_m}{2}} \prod_{i=1}^{N_m} (\det \Sigma_i)^{-0.5} \exp\left(-\frac{1}{2}\sum_{i=1}^{N_m} \underline{\delta}_i^T \Sigma_i^{-1} \underline{\delta}_i\right) \qquad (2.30)$$

The goal of the MLE is to maximize the $L\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right)$ so as to raise the probability that a measurement $\hat{\underline{z}}$ is reproduced by a certain estimated parameter vector $\hat{\underline{\theta}}$. However it is mathematically more convenient to maximize the logarithm of $L\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right)$ instead.

$$\log\left(L\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right)\right) = K - \frac{1}{2}\sum_{i=1}^{N_m} \log(\det \Sigma_i) - \frac{1}{2}\sum_{i=1}^{N_m} \underline{\delta}_i^T \Sigma_i^{-1} \underline{\delta}_i \qquad (2.31)$$

Regarding the maximization task, the constant terms in (2.31) can be left out and thus resulting in

$$\Phi\left(\hat{\underline{\theta}}\right) = -\frac{1}{2}\sum_{i=1}^{N_m} \underline{\delta}_i^T \Sigma_i^{-1} \underline{\delta}_i \rightarrow \max \qquad (2.32)$$

In general the optimization task is formulated as a minimization problem

$$\Phi\left(\hat{\underline{\theta}}\right) := -2\log\left(L\left(\hat{\underline{z}}\middle|\hat{\underline{\theta}}\right)\right) = \sum_{i=1}^{N_m} \left(\hat{\underline{z}}_i - \underline{z}_i\right)^T \Sigma_i^{-1} \left(\hat{\underline{z}}_i - \underline{z}_i\right) \rightarrow \min \qquad (2.33)$$

As a result, the objective function Φof the MLE has the structure of a LSE with Σ^{-1} as the weighting matrix. Therefore, the theory of confidence region and optimality criteria can also be applied to the nonlinear case with a crucial difference. Since the process model is nonlinear and time dependent, the estimation covariance matrix C_θ cannot be directly calculated anymore as derived in (2.15) for the linear case. Nevertheless, the lower bound of the estimation covariance matrix C_θ can be approximated by the FIM according to the asymptotic theory of MLE and the Cramer-Rao inequality (Cramér 1946; Rao 1945). It is

$$C_\theta \geq M_F^{-1} = \left(\frac{\partial \underline{z}}{\partial \underline{\theta}}^T \Sigma^{-1} \frac{\partial \underline{z}}{\partial \underline{\theta}}\right)^{-1} \qquad (2.34)$$

In case of unknown measurement covariance matrix Σ, its approximation can be used instead, according to equation (2.18)

$$\Sigma \approx \frac{1}{N_m}\sum_{i=1}^{N_m}\left(\hat{\underline{z}}_i - \underline{z}_i\right)\left(\hat{\underline{z}}_i - \underline{z}_i\right)^T \qquad (2.35)$$

The remaining unknowns are the partial derivatives terms in (2.34). They are sensitivities of the state variables $\underline{z}$ with respect to the model parameters $\underline{\theta}$ and defined as:

$$S_\theta = \left(\underline{s}_1, \cdots, \underline{s}_{N_\theta}\right) := \frac{\partial \bar{\underline{z}}}{\partial \bar{\theta}} \in R^{N_z \times N_\theta}$$
$$\underline{s}_j := \frac{\partial \underline{z}}{\partial \theta_j} = \begin{pmatrix} \frac{\partial z_1}{\partial \theta_j} \\ \vdots \\ \frac{\partial z_{N_z}}{\partial \theta_j} \end{pmatrix} \in R^{N_z \times 1}, \qquad j = 1, \dots, N_\theta \qquad (2.36)$$

Furthermore the linear-implicit form of (2.27) is assumed for the rest of the work since it generally applies to chemical engineering process models.

$$B\left(\underline{z}(t), \underline{u}(t), \underline{\theta}\right)\dot{\underline{z}}(t) = \bar{f}\left(\underline{z}(t), \underline{u}(t), \underline{\theta}, t\right), \qquad \bar{z}(t=0) = \bar{z}_0 \qquad (2.37)$$

The partial derivative with respect to the differential variable $\dot{\underline{z}}$ is usually called the mass matrix $B\left(\underline{z}(t), \underline{u}(t), \underline{\theta}\right) \in R^{N_z \times N_z}$. It is independent from $\dot{\underline{z}}$ for linear-implicit systems and might be singular in the presence of algebraic equations. To derive the parameter sensitivities S_θ, (2.37) is reformulated to

$$\underline{g}\left(\dot{\underline{z}}(t), \underline{z}(t), \underline{u}(t), \underline{\theta}, t\right) = B(\cdot)\dot{\underline{z}}(t) - \underline{f}\left(\underline{z}(t), \underline{u}(t), \underline{\theta}, t\right) = 0_{N_z \times 1} \qquad (2.38)$$

and the total differential of (2.38) is then formed with respect to $\bar{\theta}$:

$$\frac{d\underline{g}}{d\underline{\theta}} = \frac{\partial \underline{g}}{\partial \dot{\underline{z}}}\frac{\partial \dot{\underline{z}}}{\partial \underline{\theta}} + \frac{\partial \underline{g}}{\partial \underline{z}}\frac{\partial \underline{z}}{\partial \underline{\theta}} + \frac{\partial \underline{g}}{\partial \underline{\theta}} = 0_{N_z \times N_\theta} \qquad (2.39)$$

resulting in the following matrix differential equation system:

$$0_{N_z \times N_\theta} = B\frac{\partial \dot{\underline{z}}}{\partial \underline{\theta}} + \frac{\partial\left(B\dot{\underline{z}} - \underline{f}\right)}{\partial \underline{z}}\frac{\partial \underline{z}}{\partial \underline{\theta}} + \frac{\partial\left(B\dot{\underline{z}} - \underline{f}\right)}{\partial \underline{\theta}} \qquad (2.40)$$

The partial derivative with respect to the state variables is defined as J_z:

$$J_z\left(\dot{\underline{z}}(t),\underline{z}(t),\underline{u}(t),\underline{\theta},t\right) \coloneqq \left(\frac{\partial\left(B\dot{\underline{z}}\right)}{\partial\underline{z}} - \frac{\partial\underline{f}}{\partial\underline{z}}\right) \in R^{N_z\times N_z} \qquad (2.41)$$

The partial derivative with respect to the parameters is defined as J_θ:

$$J_\theta\left(\dot{\underline{z}}(t),\underline{z}(t),\underline{u}(t),\underline{\theta},t\right) \coloneqq \left(\frac{\partial\left(B\dot{\underline{z}}\right)}{\partial\underline{\theta}} - \frac{\partial\underline{f}}{\partial\underline{\theta}}\right) \in R^{N_z\times N_\theta} \qquad (2.42)$$

By using the definitions of S_θ, J_z and J_θ one can rewrite (2.40) to:

$$0_{N_z\times N_\theta} = B\dot{S}(t) + J_z S(t) + J_\theta \qquad (2.43)$$

By combining (2.27) and (2.43) one yields the extended states-sensitivities equation system DAE1, which is fundamental for solving OED tasks:

DAE1
$$\begin{aligned} 0 &= \underline{g}\left(\dot{\underline{z}}(t),\underline{z}(t),\underline{u}(t),\underline{\theta},t\right) \\ 0 &= B\dot{S}(t) + J_z S(t) + J_\theta \\ \underline{z}(t=0) &= \underline{z}_0, \qquad S_\theta(t=0) = \frac{\partial\underline{z}_0}{\partial\underline{\theta}} = 0_{N_z\times N_\theta} \end{aligned} \qquad (2.44)$$

2.4 OED–formulation as an optimal control problem

The idea to formulate the experimental design task as an optimal control problem has its roots in the pioneering work of Mehra in the 70s. The referring term to the problem was "Optimal input signals for parameter estimations" (Mehra 1974). Although the treated system is linear in the parameters the time-dependency was fully addressed which is the crucial aspect of this approach. One decade later, the treatment of the experimental design task for nonlinear dynamic models was introduced by Espie, which is formulated as DAE-systems and firstly referred to as OED (Espie & Macchietto 1989). Then, not all demands of an advanced OED task as presented in Figure 1.1 could be addressed to because of the limited possibilities of the DAE-solvers and sequential optimization algorithms from that time. State-of-the-art approaches to OED in the past decade have been successfully applied using improved sequential approaches with the single-shooting method (Barz et al. 2013; Franceschini & Macchietto 2007) and using improved sequential approaches with the multiple-shooting method (Körkel et al. 2004; Bauer et al. 2000). Recently, the paradigm of the simultaneous optimization strategy

has also reached the solution approach to OED, which is referred here as the simultaneous approach. The first application of experimental design in the sense of large scale optimization was carried out regarding precise parameter estimation and model discrimination for temperature programmed reduction experiments (Heidebrecht et al. 2011). In particular, the D-optimal criterion was used with an equidistant sampling strategy and fixed optimization end time. Nevertheless, this application example neither shows the whole strength of the simultaneous optimization approach nor gives the derivation of the related theory. The introduction to the generalized formulation and the advantages of this new approach has been recently published (Hoang et al. 2013).

2.4.1 OED – objective function

In OED, a vector of predicted response variables $\bar{y}(t) \in R^{N_y}$ is considered, whose elements usually are a subset of the state variables $\bar{z}(t)$. More generally, predicted response variables $\bar{y}(t)$ are formulated as nonlinear measurement functions of the state variables.

$$\bar{y}(t) = \bar{m}(\bar{z}(t), \bar{u}(t), \bar{\theta}, t) \qquad (2.45)$$

Their parameter sensitivities $\bar{S}_{y,\theta} \in R^{N_y \times N_\theta}$ can be calculated via the chain rule and written down directly as nonlinear function of the existing state variables $\bar{z}(t)$ and sensitivities S_θ with respect to $\bar{z}(t)$.

$$\bar{S}_{y,\theta}(t) = \begin{bmatrix} \frac{\partial m_1}{\partial z_1} & \cdots & \frac{\partial m_1}{\partial z_{N_z}} \\ \vdots & \ddots & \vdots \\ \frac{\partial m_{N_y}}{\partial z_1} & \cdots & \frac{\partial m_{N_y}}{\partial z_{N_z}} \end{bmatrix} \times [\bar{s}_1(t) \quad \cdots \quad \bar{s}_{N_\theta}(t)] = J_{m,z} \cdot S_\theta \qquad (2.46)$$

Both (2.45) and (2.46) can be straightforwardly included into the existing differential-algebraic formulation DAE1 in (2.44). For the sake of simplicity and without loss of generality, it is assumed in this work that $\bar{y}(t)$ is a subset of $\bar{z}(t)$ and in particular $N_y = N_z$. Therefore, $\bar{z}(t)$ is used when referring to the predicted response variables. Moreover, the predicted responses are generally collected at discrete points in time $t_{sp} \in R^{N_{sp}}$ according to the specific sampling points of each measured variable in the experiments.

$$\bar{z}_{i,sp}^{T} = \left[z_i\left(t_{z_1,1}\right) \quad z_i\left(t_{z_1,2}\right) \quad \cdots \quad z_i\left(t_{z_1,N_{sp,i}}\right) \right], \qquad i = 1, \dots, N_z \qquad (2.47)$$

It is assumed that the samplings of the measured variables are independent from each other. In particular the sampling time $t_{z_i,1\dots N_{sp,i}}$ and the sampling number $N_{sp,i}$ of each measured variable may be different from each other. The information content of an experiment is represented by the FIM $M_F \in R^{N_\theta \times N_\theta}$ as introduced in (2.34). For time dependent systems with discrete measurements, the definition is given in this work as follows:

$$M_F(S_{sp}) = (S_{sp})^T \cdot \Sigma_z \cdot S_{sp} \qquad (2.48)$$

Where $\Sigma_z \in R^{((N_e \cdot K)\cdot N_z)\times((N_e \cdot K)\cdot N_z)}$ denotes the measurement covariance matrix and $S_{sp} \in R^{((N_e \cdot K)\cdot N_z)\times N_\theta}$ represents the dynamic sensitivity sampling matrix, which is defined in this work as

$$S_{sp} = \begin{bmatrix} \bar{\omega}_{11} * \bar{s}_{1,11} & \bar{\omega}_{11} * \bar{s}_{2,11} & \cdots & \bar{\omega}_{11} * \bar{s}_{N_\theta,11} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{\omega}_{1K} * \bar{s}_{1,1K} & \bar{\omega}_{1K} * \bar{s}_{2,1K} & \ddots & \bar{\omega}_{1K} * \bar{s}_{N_\theta,1K} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\omega}_{N_eK} * \bar{s}_{1,1K} & \bar{\omega}_{N_eK} * \bar{s}_{2,1K} & \cdots & \bar{\omega}_{N_eK} * \bar{s}_{N_\theta,1K} \end{bmatrix} \qquad (2.49)$$

with $\bar{s}_{j,lk}$ as the discrete sensitivity vector and $\bar{\omega}_{lk}$ as the discrete sampling decision vector and $*$ the pointwise multiplication operator.

$$\bar{s}_{j,lk} := \begin{pmatrix} \left.\frac{\partial z_1}{\partial \theta_j}\right|_{lk} \\ \vdots \\ \left.\frac{\partial z_{N_z}}{\partial \theta_j}\right|_{lk} \end{pmatrix}, \qquad \bar{\omega}_{lk} := \begin{pmatrix} \omega_{1,lk} \\ \vdots \\ \omega_{N_z,lk} \end{pmatrix} \qquad (2.50)$$

$$l = 1 \dots N_e, \qquad k = 1 \dots K$$

A proper sampling strategy is crucial in OED. For this purpose, the OASE is applied to the presented simultaneous optimization approach. Each discrete sensitivity $\bar{s}_{j,lk}$ is weighted with its corresponding sampling decision vector $\bar{\omega}_{lk}$ which consists of continuous and bounded sampling decision $\omega \in R,\ 0 \leq \omega \leq 1$.

A series of optimal criteria exists in the literature for OED, which aims at maximizing system information content or minimizing the parameter correlations. The optimization task is mostly equivalent to the maximization of an appropriate measure of the FIM M_F or minimization of an appropriate measure of the covariance matrix C_θ (Franceschini & Macchietto 2008). It has been shown in (Sager 2013) based on the variational principle that if the sampling decisions ω linearly enters the Hamilton function, then, the OASE always results in a physically meaningful solution, namely 0 or 1. Regarding the formulation in (2.49), ω quadratically enters the Lagrangian, thus, it is slightly different. However, it also results in 0 or 1 because of the strict monotonicity of ω, what is discussed in section 3.5 in more detail. It is worth to mention at this point that this strict monotonicity can be guaranteed for the formulation with the FIM but not with the covariance matrix C_θ. Thus, the formulation with the FIM is strongly suggested regarding the simultaneous optimization approach. Moreover, limited sampling numbers $N_{sp,i}$ is introduced for each state so as to consider the fact that in practice the sampling numbers are much smaller than the theoretically maximum one $N_{sp} = N_e \cdot K$. Therefore, due to thetheory behind the OASE, exact $\sum_{i=1}^{N_z} N_{sp,i}$ of the total N_{sp} sampling decisions become 1 and all the others 0. The resulting optimized sampling vectors $\bar{\omega}_{lk}$ determine the optimal sampling points in (2.47). This fact has an enormous impact on the optimization algorithm since a much more complicated MINLP formulation can be avoided. The application of the OASE is discussed in chapter 0 in more detail.

2.4.2 Formulation with sequential optimization approaches

Although the differences between the sequential and the simultaneous optimization strategies are crucial, all of them have the extended state sensitivity system DAE1 in common. However, the sequential approach additionally needs sensitivities with respect to the controls S_u and higher order sensitivities $S_{\theta u}$ and $S_{\theta uu}$which are given by DAE2.

DAE2

$$\begin{aligned} \dot{S}_u &= J_z S_u + J_u \in R^{N_z \times N_u} \\ \dot{S}_{\theta u} &= \underline{f}_{\theta u}\left(S_{\theta u}, S_\theta, \underline{z}(t), \underline{u}(t), \underline{\theta}, t\right) \in R^{(N_z \cdot N_\theta) \times N_u} \\ \dot{S}_{\theta uu} &= \underline{f}_{\theta uu}\left(S_{\theta uu}, S_{\theta u}, S_\theta, \underline{z}(t), \underline{u}(t), \underline{\theta}, t\right) \in R^{(N_z \cdot N_\theta) \times N_u \times N_u} \\ S_u(t=0) &= 0_{N_z \times N_u}, \qquad S_{\theta u}(t=0) = 0_{(N_z \cdot N_\theta) \times N_u}, \\ S_{\theta uu}(t=0) &= 0_{(N_z \cdot N_\theta) \times N_u \times N_u} \end{aligned} \qquad (2.51)$$

Moreover, if the initial conditions of the state variables are also optimized then the differential equation system with respect to the initial conditions DAE2$_0$ has also to be included into the optimization formulation.

DAE2$_0$

$$\frac{dS_0}{dt} = J_z S_{z_0} \in R^{N_z \times N_u}$$
$$S_0(t=0) = 1_{N_z \times N_z}, \qquad S_0 := \frac{\partial \underline{z}}{\partial \underline{z}_0} \tag{2.52}$$

As a result three DAE-systems have to be integrated during each optimization step. The resulting sequential optimization problem is formulated as

OP1

$$\min_{\underline{u}_d, \underline{\omega}} \Phi\left(V_\theta\left(S_\theta, \underline{\omega}\right)\right) \text{ or } \max_{\underline{u}_d, \underline{\omega}} \Phi\left(M_F\left(S_\theta, \underline{\omega}\right)\right) \tag{2.53}$$

Subject to:

$$\text{DAE1} \cup \text{DAE2} \cup \text{DAE2}_0 \tag{2.54}$$
$$\underline{u}_L \le \underline{u}_d \le \underline{u}_U, \qquad e = 1 \dots N_d$$

2.4.3 Formulation with simultaneous optimization approaches

By stacking the columns of S_θ as defined in (2.36), (2.43) is reformulated to a form, which is more suitable for the simultaneous optimization approach

$$0_{N_z \cdot N_\theta \times 1} = \bar{g}_\theta = \begin{pmatrix} \bar{g}_{\theta_1} \\ \vdots \\ \bar{g}_{\theta_{N_\theta}} \end{pmatrix} := \begin{pmatrix} B \cdot \dot{\bar{s}}_1 \\ \vdots \\ B \cdot \dot{\bar{s}}_{N_\theta} \end{pmatrix} + \begin{pmatrix} J_z \cdot \bar{s}_1 \\ \vdots \\ J_z \cdot \bar{s}_{N_\theta} \end{pmatrix} + \begin{pmatrix} J_{\theta_1} \\ \vdots \\ J_{\theta_{N_\theta}} \end{pmatrix} \tag{2.55}$$

with

$$\begin{pmatrix} \bar{f}_{\theta_1} \\ \vdots \\ \bar{f}_{\theta_{N_\theta}} \end{pmatrix} := \begin{pmatrix} J_z \cdot \bar{s}_1 \\ \vdots \\ J_z \cdot \bar{s}_{N_\theta} \end{pmatrix} + \begin{pmatrix} J_{\theta_1} \\ \vdots \\ J_{\theta_{N_\theta}} \end{pmatrix} \tag{2.56}$$

Now (2.44) can be rewritten to:

DAE3

$$0_{N_z(1+N_\theta)\times 1} = \begin{bmatrix} \bar{g}(\dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t) \\ \bar{g}_\theta\left(\dot{\bar{s}}_1, \dots, \dot{\bar{s}}_{N_\theta}, \bar{s}_1, \dots, \bar{s}_{N_\theta}, \dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t\right) \end{bmatrix}$$
$$:= \bar{c}\left(\dot{\bar{z}}(t), \bar{z}(t), \dot{\bar{s}}_1, \dots, \dot{\bar{s}}_{N_\theta}, \bar{s}_1, \dots, \bar{s}_{N_\theta}, \bar{u}(t), \bar{\theta}, t\right) \tag{2.57}$$
$$\bar{z}(t=0) = \bar{z}_0\,, \qquad \bar{s}_j(t=0) = \frac{\partial \bar{z}_0}{\partial \theta_j}, \qquad j = 1, \dots, N_\theta$$

By converting all time-dependent functions of DAE3 into discrete variables the standard OED task can be formulated as a NLP optimization problem as follows:

OP2

$$\min_{\bar{z}_{lk},S_{\theta,lk},u_{q,lk_q},h_l,\bar{\omega}_{lk}} \Phi\left(C_\theta\left(S_{sp}\right)\right)$$

$$\text{or} \quad \max_{\bar{z}_{lk},S_{\theta,lk},u_{q,lk_q},h_l,\bar{\omega}_{lk}} \Phi\left(M_F\left(S_{sp}\right)\right)$$

Subject to:

$$\begin{gathered}
0 = \bar{c}\left(\bar{z}_{lk}, S_{lk}, u_{q,lk_q}, \bar{\theta}, \tau_k, h_l\right) \\
l = 1 \dots N_e, \qquad k = 1 \dots K+1, \qquad k_q = 1 \dots K_{u,q}, \\
q = 1 \dots N_u \\
\bar{z}_{11} = \bar{z}_0, \qquad S_{\theta 11} = \frac{\partial \bar{z}_0}{\partial \bar{\theta}} \\
\bar{z}_{l+1,k=1} = \bar{z}_{l,k=K+1}, \qquad S_{\theta_{l+1,k=1}} = S_{\theta_{l+1,k=K+1}} \\
l = 1, \dots, N_e - 1 \\
\sum_{l=1}^{N_e}\sum_{k=1}^{K} \omega_{i,lk} = N_{sp,i}, \qquad 0 \le \omega_k \le 1, \qquad i = 1, \dots, N_z \\
\sum_{l=1}^{N_e} h_l = t_f, \qquad t_f \le t_{max}, \qquad h_L \le h_l \le h_U \\
\bar{z}_{0,L} \le \bar{z}_{11} \le \bar{z}_{0,U}, \qquad \bar{z}_L \le \bar{z}_{lk} \le \bar{z}_U, \qquad \bar{u}_L \le \bar{u}_{lk} \le \bar{u}_U
\end{gathered} \tag{2.58}$$

The crucial advantage of the simultaneous formulation in OP2 in comparison with the sequential formulation of OP1 is that the entire sensitivity systems DAE2 and $DAE2_0$ can be taken out since the equality constraints can be directly differentiate with respect to all variables. If the problem contains many parameters, state and control variables, one can easily see that an enormous amount of computational effort can be saved. Furthermore the simultaneous formulation allows a flexible and convenient way to use high order and continuous control trajectories. Since the control functions are also discretized with the OCFEM, one is free to choose different orders K_c for each control variable u_c. Furthermore, the handling of the optimization of the initial states is straightforward since they appear explicitly in OP2. Therefore, they are treated in the same manner as the discretized dynamic controls. Inequality conditions for state and control variables can be directly embedded into the formulation over the whole domain, which is superior in comparison to the sequential approach.

Scaling

The FIM strongly depends on a scaling of the parameter sensitivities. Since its formulation is based on the absolute parameter values, the influence of parameters with small values is much bigger than those with high values. Therefore, parameter sensitivities have to be scaled so as to take into account the dimension gap of their different nature. A common way to scale sensitivities regarding a parameter θ_i is to multiply with the parameter itself (Franceschini & Macchietto 2008):

$$\tilde{s}_j := \theta_j \bar{s}_j, \qquad j = 1, \dots, N_\theta \tag{2.59}$$

Thus, the scaled FIM is formulated as:

$$M_F\left(S_{sp}(\tilde{s})\right) = \left(S_{sp}(\tilde{s})\right)^T \cdot \Sigma_z \cdot S_{sp}(\tilde{s}) \tag{2.60}$$

For the sake of simplicity, the scaled FIM is also referred as M_F. The reader is asked to keep in mind that all the following equations and calculations are based on scaled sensitivities. According to Table 7, the used objective functions (A, D, E) are formulated here with respect to the Fischer Information matrix

Table 8 Experimental design criteria – FIM

A-criterion	$\text{trace}\, M_F \rightarrow \max$
D-criterion	$\det M_F \rightarrow \max$
E-criterion	$\lambda_{\min} \rightarrow \max$

It is worth to emphasize that the formulation of the objective function with respect to the covariance matrix C_θ as shown in Table 7 does not guarantee a physical meaningful sampling strategy when including the OASE (see 3.5 in more detail). Therefore OP2 is only used with respect to the FIM. The optimization variables of OP2 are summarized in the vector $\bar{x}$ for the following derivation.

$$\begin{aligned} \bar{x}^T := \big[& \bar{z}_{11}^T \quad \cdots \quad \bar{z}_{N_e K+1}^T \quad \bar{s}_{1,11}^T \quad \cdots \quad \bar{s}_{N_\theta, N_e K+1}^T \quad u_{1,11} \quad \cdots \\ & u_{N_u, N_e K_{N_u}} \quad \bar{z}_0^T \quad h_1 \quad \cdots \quad h_{N_e} \quad \bar{\omega}_{11}^T \quad \cdots \quad \bar{\omega}_{N_e K}^T \big] \end{aligned} \tag{2.61}$$

The derivatives of the optimal criteria are developed via the chain rule

$$\frac{\partial \Phi\left(\mathrm{M_F}\left(\mathrm{S_{sp}}\right)\right)}{\partial \bar{\mathrm{x}}} = \frac{\partial \Phi(\mathrm{M_F})}{\partial \mathrm{M_F}} \cdot \frac{\partial \mathrm{M_F}\left(\mathrm{S_{sp}}\right)}{\partial \bar{\mathrm{x}}} \tag{2.62}$$

The derivatives of the A, D, and E-criteria only differ from each other in the first term of the product in (2.62) which is $\in \mathrm{R}^{1\times(\mathrm{N_\theta}\cdot\mathrm{N_\theta})}$:

$$\frac{\partial \Phi_{\mathrm{A}}(\mathrm{M_F})}{\partial \mathrm{M_F}} = \frac{\partial\left(\mathrm{tr}(\mathrm{M_F})\right)}{\partial \mathrm{M_F}} = \left(\mathrm{vec}\left(\mathrm{I_{N_\theta}}\right)\right)^{\mathrm{T}} \tag{2.63}$$

$$\frac{\partial \Phi_{\mathrm{D}}(\mathrm{M_F})}{\partial \mathrm{M_F}} = \frac{\partial\left(\det(\mathrm{M_F})\right)}{\partial \mathrm{M_F}} = \det(\mathrm{M_F}) \cdot \left(\mathrm{vec}\left(\mathrm{M_F^{-T}}\right)\right)^{\mathrm{T}} \tag{2.64}$$

$$\begin{aligned}\frac{\partial \Phi_{\mathrm{E}}(\mathrm{M_F})}{\partial \mathrm{M_F}} &= \frac{\partial\left(\lambda_{\min}(\mathrm{M_F})\right)}{\partial \mathrm{M_F}} \\ &= \left(\mathrm{v_{\lambda_{min}}}^{\mathrm{T}} \cdot \mathrm{v_{\lambda_{min}}}\right)^{-1} \cdot \mathrm{v_{\lambda_{min}}}^{\mathrm{T}} \cdot \left(\mathrm{I_{N_\theta}} \otimes \mathrm{v_{\lambda_{min}}}^{\mathrm{T}}\right)\end{aligned} \tag{2.65}$$

Here $\mathrm{v_{\lambda_{min}}}$ denotes the eigenvector corresponding to the smallest eigenvalue of $\mathrm{M_F}$. For (2.64), the following matrix calculus rule is used. The derivative of the determinant of the matrix $\mathrm{A} \in \mathrm{R}^{\mathrm{n}\times\mathrm{n}}$ with respect to its elements is given as

$$\begin{gathered}\frac{\mathrm{d}|\mathrm{A}|}{\mathrm{dA}} = \frac{\mathrm{d}|\mathrm{A}|}{\mathrm{d}\left(\mathrm{vec}(\mathrm{A})\right)} = |\mathrm{A}|\left(\mathrm{vec}(\mathrm{A^{-T}})\right)^{\mathrm{T}} \in \mathrm{R}^{1\times\mathrm{n}\cdot\mathrm{n}} \\ \mathrm{A^{-T}} = (\mathrm{A^{-1}})^{\mathrm{T}}\end{gathered} \tag{2.66}$$

The second term of (2.62) represents the Jacobian of $\mathrm{M_F}$ with respect to the elements of X, which is the same for all criteria.

$$\begin{gathered}\frac{\partial \mathrm{M_F}}{\partial \bar{\mathrm{x}}} = \frac{\partial\left(\mathrm{vec}(\mathrm{M_F})\right)}{\partial \bar{\mathrm{x}}} \in \mathrm{R}^{\mathrm{N_\theta}\cdot\mathrm{N_\theta}\times\left(\mathrm{N_z}\cdot\mathrm{N_K}+\mathrm{N_z}\cdot\mathrm{N_\theta}\cdot\mathrm{N_K}+\mathrm{N_{K_u}}+\mathrm{N_e}+\mathrm{N_\omega}\right)} \\ \mathrm{N_K} = \mathrm{N_e}\cdot(\mathrm{K}+1), \mathrm{N_{K_u}} = \sum_{\mathrm{q}=1}^{\mathrm{N_u}} \mathrm{N_e}\cdot\left(\mathrm{K_{u,q}}+1\right), \mathrm{N_\omega} = \sum_{\mathrm{i}=1}^{\mathrm{N_z}} \mathrm{N_e}\cdot\mathrm{K}\end{gathered} \tag{2.67}$$

In order to obtain the second-order derivatives analytically, it is required to use the Kronecker and box matrix products (Olsen et al. 2012). To avoid the more complicated derivation, symbolic differentiation can be used to directly calculate the second

derivatives. For this purpose, the results from equation (2.62) have to be differentiated one more time with respect to $\bar{x}$.

2.5 Exploiting the structure of the constraint derivatives

For the sake of clarity, the derivation of the constraints derivatives is applied on Equation (2.57) with respect to the continuous optimization variables $\left[\bar{z}(t), \bar{s}_1(t), \dots, \bar{s}_{N_\theta}(t), \bar{u}(t)\right]$. Since only the objective function depends on $\bar{\omega}$ but not the constraints, all constraint derivatives with respect to $\bar{\omega}$ become zero and will not be additionally included in the following derivation.

In order to obtain the derivatives with respect to the fully discretized optimization variables $\left[\bar{z}_{lk}, \bar{s}_{1,lk}, \dots, \bar{s}_{N_\theta,lk}, \bar{z}_0, \bar{u}_{lc}, h_l\right]$, one more derivation step over the discretization equations (3.3) to (3.5) is needed which is straightforward and not be further discussed. The important idea here is the information about the extremely sparse structure of the derivatives and the fact that most parts of them can be reused from previous calculations, which is illustrated in the following derivation. The optimization variables for the derivatives are represented by

$$\bar{x} := \begin{pmatrix} \bar{z} \\ \bar{s}_1 \\ \vdots \\ \bar{s}_{N_\theta} \\ \bar{u} \end{pmatrix} \in R^{(N_z + N_z \cdot N_\theta + N_u) \times 1} \qquad (2.68)$$

2.5.1 First-order constraint derivatives

The first-order derivative of the constraints with respect to the optimization variables $\bar{x}$ is then

$$\frac{\partial \bar{c}}{\partial \bar{x}} = \frac{\partial}{\partial \bar{x}} \begin{pmatrix} \bar{g} \\ \bar{g}_\theta \end{pmatrix} = \begin{bmatrix} J_z & 0 & 0 & \cdots & 0 & J_u \\ J_z^{\theta_1} & J_s^{\theta_1} & 0 & \vdots & 0 & J_u^{\theta_1} \\ J_z^{\theta_2} & 0 & J_s^{\theta_{N_2}} & \ddots & 0 & J_u^{\theta_2} \\ \vdots & \vdots & \cdots & \ddots & \vdots & \vdots \\ \underbrace{J_z^{\theta_{N_\theta}}}_{\text{w.r.t. } \bar{z}} & \underbrace{0 \quad \cdots \quad \cdots \quad J_s^{\theta_{N_\theta}}}_{\text{w.r.t. } \bar{s}} & & & & \underbrace{J_u^{\theta_{N_\theta}}}_{\text{w.r.t. } \bar{u}} \end{bmatrix} \qquad (2.69)$$

$$\frac{\partial \bar{c}}{\partial \bar{x}} \in R^{(N_z + N_z \cdot N_\theta) \times (N_z + N_z \cdot N_\theta + N_u)}$$

The first-order derivative of the system equations $\bar{g}$ with respect to the state variables $\bar{z}$ is the Jacobian J_z defined in (2.41), which has already been calculated to formulate the sensitivity equations$\bar{g}_\theta$, and thus, it can be fully reused. The derivatives of $\bar{g}$ regarding the control variables $\bar{u}$ are represented by

$$J_u(\dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t) := \frac{\partial \bar{g}}{\partial \bar{u}} = \left(\frac{\partial \left(B\dot{\bar{z}}(t) \right)}{\partial \bar{u}} - \frac{\partial \bar{f}}{\partial \bar{u}} \right) \qquad (2.70)$$

$$J_u \in R^{N_z \times N_u}$$

Furthermore, the derivative of the j-th part of $\bar{f}_\theta$ in (2.56) with respect to the sensitivities $\bar{s}_j$ is also J_z since $\bar{s}_j$ is linear in $\bar{f}_{\theta_j}$ leading to:

$$J_s^{\theta_j}(\dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t) := \frac{\partial \bar{g}_{\theta_j}}{\partial \bar{s}_j} = \left(\frac{\partial \left(B\dot{\bar{s}}_j \right)}{\partial \bar{s}_j} + J_z \right) \qquad (2.71)$$

$$J_s^{\theta_j} \in R^{N_z \times N_z}$$

The derivatives of $\bar{g}_{\theta_j}$ with respect to $\bar{z}$ are defined as

$$J_z^{\theta_j} := \frac{\partial \bar{g}_{\theta_j}}{\partial \bar{z}} = \frac{\partial}{\partial \bar{z}} \left[B \cdot \dot{\bar{s}}_j + J_z \cdot \bar{s}_j + \frac{\partial B}{\partial \theta_j} \dot{\bar{z}}(t) - \frac{\partial \bar{f}}{\partial \theta_j} \right]$$

$$= \left[\frac{\partial \left(B\dot{\bar{s}}_j \right)}{\partial \bar{z}} + \left[\frac{\partial J_z}{\partial z_1} \bar{s}_j \quad \cdots \quad \frac{\partial J_z}{\partial z_{N_z}} \bar{s}_j \right] + \frac{\partial}{\partial \bar{z}} \left(\frac{\partial B}{\partial \theta_j} \dot{\bar{z}}(t) \right) - \frac{\partial}{\partial \bar{z}} \left(\frac{\partial \bar{f}}{\partial \theta_j} \right) \right] \qquad (2.72)$$

$$J_z^{\theta_j} \in R^{N_z \times N_z}, \qquad j = 1 \dots N_\theta$$

And in the same way with respect to $\bar{u}$

$$J_u^{\theta_j} := \frac{\partial \bar{g}_{\theta_j}}{\partial \bar{u}} = \frac{\partial}{\partial \bar{u}} \left[B \cdot \dot{\bar{s}}_j + J_z \cdot \bar{s}_j + \frac{\partial B}{\partial \theta_j} \dot{\bar{z}}(t) - \frac{\partial \bar{f}}{\partial \theta_j} \right]$$

$$= \left[\frac{\partial \left(B\dot{\bar{s}}_j \right)}{\partial \bar{u}} + \left[\frac{\partial J_z}{\partial u_1} \bar{s}_j \quad \cdots \quad \frac{\partial J_z}{\partial u_{N_u}} \bar{s}_j \right] + \frac{\partial}{\partial \bar{u}} \left(\frac{\partial B}{\partial \theta_j} \dot{\bar{z}}(t) \right) - \frac{\partial}{\partial \bar{u}} \left(\frac{\partial \bar{f}}{\partial \theta_j} \right) \right] \qquad (2.73)$$

$$J_u^{\theta_j} \in R^{N_z \times N_u}, \qquad j = 1 \dots N_\theta$$

2.5.2 Second order constraint derivatives

Based on the derivation of the second order constraint derivatives, one can conclude that most of the derived partial Jacobian can be reused. Therefore, many of the derivative calculation steps can be reduced. The first part of the second order constraints derivatives can be achieved by differentiating the system equations $\bar{g}$ as part of $\bar{c}$ in (2.57). For the i-th part g^i, the following Hessian can be formulated:

$$H_{g^i} := \begin{bmatrix} J_{zz}^i & 0_{N_z \times N_z \cdot N_\theta} & J_{zu}^i \\ 0_{N_z \cdot N_\theta \times N_z} & 0_{N_z \cdot N_\theta \times N_z \cdot N_\theta} & 0_{N_z \cdot N_\theta \times N_u} \\ J_{uz}^i & 0_{N_u \times N_z \cdot N_\theta} & J_{uu}^i \end{bmatrix} \tag{2.74}$$

$$H_{g^i} \in R^{(N_z + N_z \cdot N_\theta + N_u) \times (N_z + N_z \cdot N_\theta + N_u)}$$

$$i = 1 \dots N_z$$

With

$$J_{zz}^i := \left(\frac{\partial^2 \left(B \dot{\bar{z}}(t) \right)^i}{\partial \bar{z}^2} - \frac{\partial^2 f^i}{\partial \bar{z}^2} \right) \in R^{N_z \times N_z} \tag{2.75}$$

and

$$J_{zu}^i = \left(\frac{\partial^2 \left(B \dot{\bar{z}}(t) \right)^i}{\partial \bar{z} \, \partial \bar{u}} - \frac{\partial^2 f^i}{\partial \bar{z} \, \partial \bar{u}} \right) \in R^{N_z \times N_u} \tag{2.76}$$

$$J_{uz}^i = J_{zu}^i$$

where the i-th row of a matrix or the i-th entry of a column is written shortly as $(\cdot)^i$. The partial Hessian $J_{zz,i}$ can be assembled from the i-th rows of the terms $\partial J_z / \partial z_1 \dots \partial J_z / \partial z_{N_z}$ in (2.72). The same applies for the partial Hessian $J_{zu,i}$ regarding the terms $\partial J_z / \partial u_1 \dots \partial J_z / \partial u_{N_u}$ in (2.73). Only $J_{uu,i}$ has to be calculated additionally. Furthermore, the structure of H_{g^i} is extremely sparse regarding its dimension, which can in fact be favorably integrated to advanced optimization routines which can take advantages of sparsity of the system. In analogy to obtaining the partial Hessians H_{g^i} for the i-th row

of $\bar{g}$, one also obtains the partial Hessians $H_{g_{\theta_j}^i}$ for the i-th part of each partial sensitivity equations$\bar{g}_{\theta_i}$:

$$H_{g_{\theta_j}^i} := \begin{bmatrix} J_{zz,i}^{\theta_j} & 0_{N_z\times(j-1)\cdot N_z} & J_{zs_j,k}^{\theta_j} & 0_{N_z\times(N_\theta-j+1)\cdot N_z} & J_{zu,i}^{\theta_j} \\ 0_{(j-1)\cdot N_z\times N_z} & \vdots & \cdots & \vdots & 0_{(j-1)\cdot N_z\times N_u} \\ J_{s_jz,k}^{\theta_j} & \vdots & 0_{N_z\times N_\theta} & \vdots & J_{s_ju,i}^{\theta_j} \\ 0_{(N_\theta-j+1)\cdot N_z\times N_z} & \vdots & \cdots & \vdots & 0_{(N_\theta-j+1)\cdot N_z\times N_u} \\ J_{uz,i}^{\theta_j} & 0_{N_u\times(j-1)\cdot N_z} & J_{us_j,i}^{\theta_j} & 0_{N_u\times(N_\theta-j+1)\cdot N_z} & J_{uu,i}^{\theta_j} \end{bmatrix} \quad (2.77)$$

$$H_{g_{\theta_j}^i} \in R^{(N_z+N_z\cdot N_\theta+N_u)\times(N_z+N_z\cdot N_\theta+N_u)}$$

$$i = 1 \dots N_z, \qquad j = 1 \dots N_\theta$$

Equation (2.77) reveals that the structure of $H_{g_{\theta_j}^i}$ is also extremely sparse.The partial Hessians regarding the states are:

$$J_{zz,i}^{\theta_j} := \frac{\partial\left(J_z^{\theta_j}\right)^i}{\partial\bar{z}} = \frac{\partial^2\left(J_z\cdot\bar{s}_j\right)^i}{\partial\bar{z}^2} + \frac{\partial^2}{\partial\bar{z}^2}\left(\frac{\partial B}{\partial\theta_j}\dot{\bar{z}}(t)\right)^i - \frac{\partial^2}{\partial\bar{z}^2}\left(\frac{\partial f^i}{\partial\theta_j}\right) \quad (2.78)$$

$$J_{zz,j}^{\theta_j} \in R^{N_z\times N_z}$$

and the partial Hessians regarding states and sensitivities:

$$J_{sz,i}^{\theta_j} := \frac{\partial\left(J_s^{\theta_j}\right)^i}{\partial\bar{z}} = \frac{\partial}{\partial\bar{z}}\left(\frac{\partial\left(B\dot{\bar{s}}_j\right)}{\partial\bar{s}_j}\right)^i + J_{zz}^i \quad (2.79)$$

$$J_{sz,i}^{\theta_j} \in R^{N_z\times N_z\cdot N_\theta}, \qquad J_{zs,i}^{\theta_j} = J_{sz,i}^{\theta_j}$$

The terms $J_{zz,i}$ in (2.79) can be directly carried over from (2.75). It is analogous to the partial hessian regarding controls and sensitivities:

$$J_{su,i}^{\theta_j} := \frac{\partial\left(J_s^{\theta_j}\right)^i}{\partial\bar{u}} = \frac{\partial}{\partial\bar{u}}\left(\frac{\partial\left(B\dot{\bar{s}}_j\right)}{\partial\bar{s}_j}\right)^i + J_{zu}^i \quad (2.80)$$

$$J_{su,i}^{\theta_j} \in R^{N_z\times N_u}, \qquad J_{us,i}^{\theta_j} = J_{su,i}^{\theta_j}$$

where the terms J_{zu}^{i} in (2.80) can be directly carried over from (2.76). Furthermore, the partial Hessians regarding the controls are:

$$J_{uu,i}^{\theta_j} := \frac{\partial \left(J_s^{\theta_j}\right)^i}{\partial \bar{u}}$$
$$= \frac{\partial^2 \left(B \cdot \dot{\bar{s}}_j\right)^i}{\partial \bar{u}^2} + \frac{\partial^2 \left(J_z \cdot \bar{s}_j\right)^i}{\partial \bar{u}^2} + \frac{\partial^2}{\partial \bar{u}^2}\left(\frac{\partial B}{\partial \theta_j}\dot{\bar{z}}(t)\right)^i \qquad (2.81)$$
$$- \frac{\partial^2}{\partial \bar{u}^2}\left(\frac{\partial f^i}{\partial \theta_j}\right)$$
$$J_{uu,i}^{\theta_j} \in R^{N_u \times N_u}$$

and with respect to the controls and states:

$$J_{uz,i}^{\theta_j} := \frac{\partial \left(J_u^{\theta_j}\right)^i}{\partial \bar{z}}$$
$$= \frac{\partial^2 \left(B \cdot \dot{\bar{s}}_j\right)^i}{\partial \bar{u}\,\partial \bar{z}} + \frac{\partial^2 \left(J_z \cdot \bar{s}_j\right)^i}{\partial \bar{u}\,\partial \bar{z}} + \frac{\partial^2}{\partial \bar{u}\,\partial \bar{z}}\left(\frac{\partial B}{\partial \theta_j}\dot{\bar{z}}(t)\right)^i \qquad (2.82)$$
$$- \frac{\partial^2}{\partial \bar{u}\,\partial \bar{z}}\left(\frac{\partial f^i}{\partial \theta_j}\right)$$
$$J_{uz,i}^{\theta_j} \in R^{N_u \times N_z}, \qquad J_{zu,i}^{\theta_j} = J_{uz,i}^{\theta_j}$$

3 Simultaneous optimization approach to OED

The fundamental idea of the simultaneous optimization approach to OED is the total discretization of the dynamic process model. Therefore the applied discretization method fundamentally influences the outcome of the optimization steps. A very detailed comparison between well-known discretization methods like finite difference method (FDM), Galerkin finite element method (GFEM), orthogonal collocation method (OCM) and orthogonal collocation on finite element method (OCFEM) is given in the standard book about numerical methods in chemical engineering by Finlayson (Finlayson 1980). The comparison reveals that the OCM has the best characteristic among all approaches; in particular it gives best accuracy with the fewest total discretization points. However, it fails in cases of process variables with steep profiles. Therefore, the OCFEM is recommended so as to cover a broader application field (Carey & Finlayson 1975). Basically, the OCFEM can also be applied to the discretization in space covering PDE systems. However, if the process contains steep profiles along the spatial coordinates a mixed discretization is highly recommended including OCFEM for the time and upwind type GFEM for the spatial discretization (Brooks & Hughes 1982). Without loss of generality only DAE systems are discussed in this thesis.

3.1 Concept of the OCFEM

Considering following linear implicit DAE-system as introduced in (2.37):

$$B(\bar{z}(t), \bar{u}(t), \bar{\theta}, t) \cdot \dot{\bar{z}}(t) = \bar{f}(\bar{z}(t), \bar{u}(t), \bar{\theta}, t), \qquad \bar{z}(t = 0) = \bar{z}_0 \qquad (3.1)$$

The pure algebraic equations of the DAE-system are taken into account with a singular mass matrix B, and thus, they are not written explicitly. All time dependencies of the state and control variables are discretized via the OCFEM. Let the number of finite elements be N_e, the index of each element l and the index of the collocation points in each element k, then the discretization scheme of a state variable in one element can be illustrated in Figure 3.1.

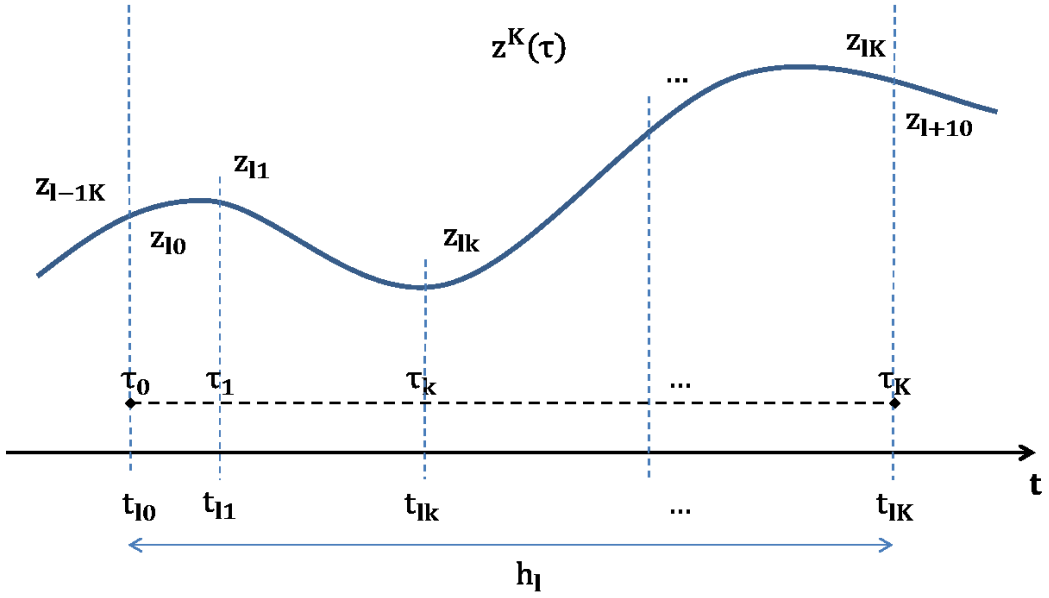


Figure 3.1 Discretization scheme of state variables

The functional section in each element is approximated by polynomials of order K. Within each element the overall discrete time t_{lk} is projected to the local normalized collocation time τ_k.

$$t_{lk} = t_{l-1K} + \tau_k h_l, \qquad \tau_k \in [0,1]$$
$$l = 1, \dots, N_e \quad , \qquad k = 0, \dots, K \quad , \quad h_0 = t_0 \tag{3.2}$$

Lagrange-polynomials are chosen as base functions. The advantage of the Lagrange polynomials is that the coefficients directly represent the desired state values, what simplifies the evaluation of the results. The function of the state variables in element l is represented by the Lagrange polynomials $\varphi_{lk}(t)$:

$$\underline{z}_l^K(t) = \sum_{k=0}^{K} \underline{z}_{lk} \varphi_{lk}(t)$$
$$\varphi_{lk}(t) = \prod_{\substack{i=0 \\ i \neq k}}^{K} \frac{(t - t_{li})}{(t_{lk} - t_{li})} \tag{3.3}$$

The time dependent control variables are discretized analogously but using polynomial approximations of possibly different control orders $K_{u,q}$.

$$u_{q,l}^{K_{u,q}}(t) = \sum_{k=0}^{K_{u,q}} u_{q,lk}\psi_{lk}(t), \qquad q = 1 \dots N_u$$

$$\psi_{lk}(t) = \prod_{\substack{i=0 \\ i \neq k}}^{K_{u,q}} \frac{(t - t_{li})}{(t_{lk} - t_{li})}, \tag{3.4}$$

One main advantage of this approach is that the control orders can be chosen totally independent from each other. It is also independent from the discretization of the state variables leading to highly flexible formulation possibilities. The discretization schemes for control variables of first and zero-orders are exemplary shown in Figure 3.2 and Figure 3.3.

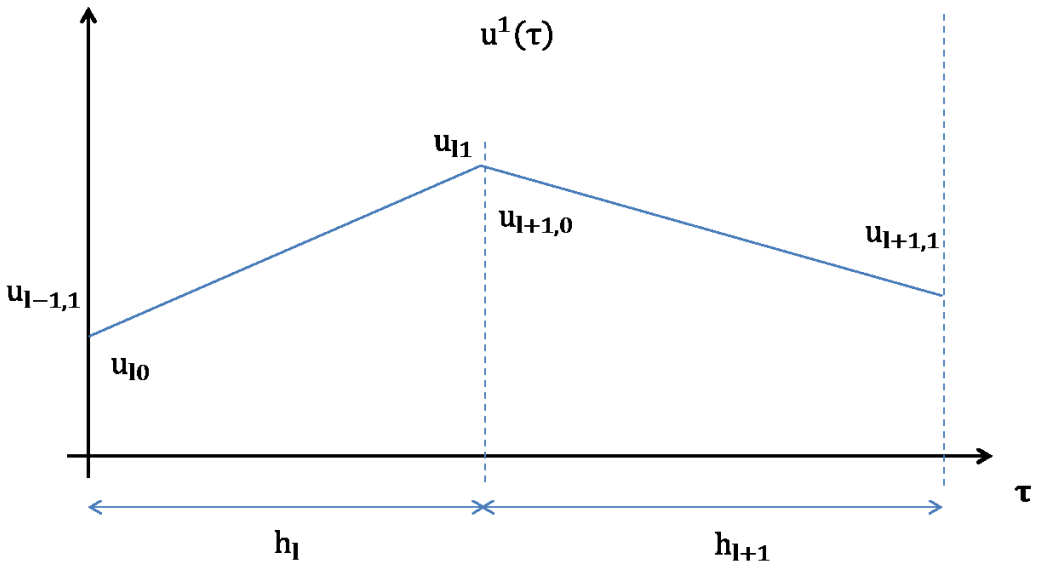


Figure 3.2 Discretization scheme of first-order control variables

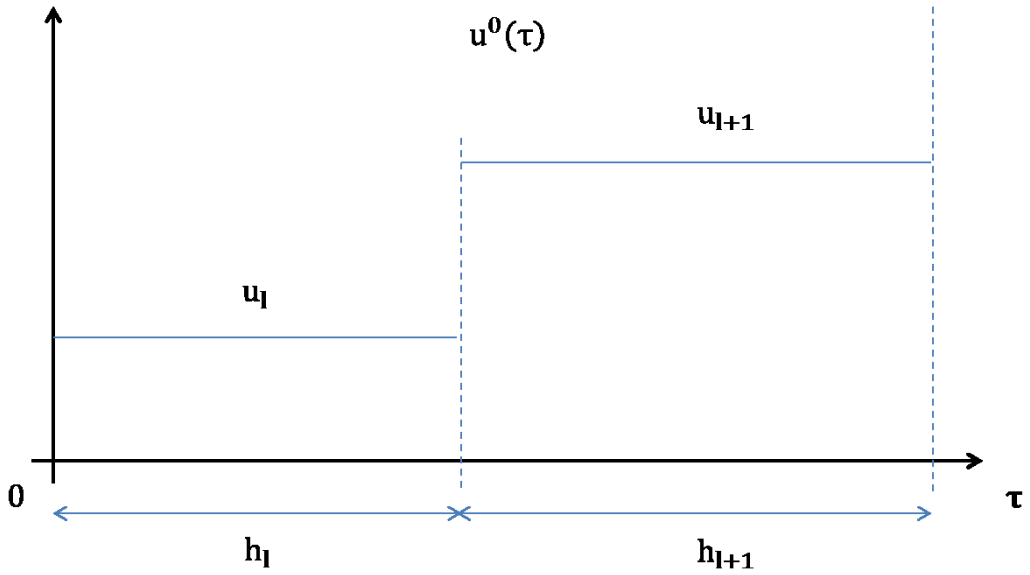


Figure 3.3 Discretization scheme of zero-order control variables

Since the discretized functions of the state variables are only time dependent in the base functions, the time derivatives of state variables can be written as:

$$\dot{\underline{z}}_l(t) = \sum_{k=0}^{K} \underline{z}_{lk} \dot{\varphi}_{lk}(t) \qquad (3.5)$$

By applying the chain rule on (3.2) for every $\dot{\varphi}_{lk}(t)$, the following simplification can be obtained:

$$\dot{\varphi}_{lk}(t) = \frac{d\varphi_{lk}(t)}{dt} = \frac{\partial \varphi_{lk}(\tau)}{\partial \tau}\frac{\partial \tau}{\partial t} = \frac{\partial \varphi_{lk}(\tau)}{\partial \tau}\frac{1}{h_l} \qquad (3.6)$$
$$l = 1, \dots, N_e \quad ; \quad k = 0, \dots, K$$

Since the same collocation method and the same polynomial order are applied to all elements of the state variables, one gets

$$\varphi_{lk}(t) = \varphi_k(\tau) = \prod_{\substack{i=0 \\ i \neq k}}^{K} \frac{(\tau - \tau_i)}{(\tau_k - \tau_i)} \qquad (3.7)$$

$$\frac{\partial \varphi_{lk}(\tau)}{\partial \tau} = \frac{\partial \varphi_k(\tau)}{\partial \tau} \qquad (3.8)$$

Thus, equation (3.5) can be rewritten as:

$$\dot{\underline{z}}_l(t_{lk}) = \sum_{k=0}^{K} \underline{z}_{lk} \frac{1}{h_l} \frac{\partial \varphi_k(\tau)}{\partial \tau} \qquad (3.9)$$

Furthermore, continuity conditions ensure that no discontinuities arise between the elements of the discretized functions. They are generally formulated for the OCFEM as Legendre-type:

$$\underline{\tilde{z}}_{l+1,k=0} = \sum_{k=0}^{K} \underline{\tilde{z}}_{lk} \varphi_k(\tau = 1) \quad , \qquad l = 1, \dots, N_e - 1 \qquad (3.10)$$

where $\underline{\tilde{z}}$ represents the system state variables, which only belong to the differential equations in (3.1). The meaning of (3.10) is that the polynomial function $\tilde{z}_l(t)$ of the l-th element is extrapolated to its end point and specifies thereby the initial value for the subsequent $(l+1)$-th element. For the Radau-type OCFEM, the continuity conditions are:

$$\begin{gathered} \underline{\tilde{z}}_{l+1,j=0} = \underline{\tilde{z}}_{l,j=K} \\ l = 1, \dots, N_e - 1 \end{gathered} \qquad (3.11)$$

This collocation strategy was applied in (Kameswaran & Biegler 2008) and it was pointed out as superior with respect to solving NLP problems since it directly provides the values of the element's end point and is the OCFEM with the second highest precision after the Legendre-type. Applying the Radau-OCFEM on (3.1), one gets:

$$\sum_{k=0}^{K} B\left(\underline{z}_{lk}, u_{lk_u}, \bar{\theta}\right) \underline{z}_{lk} \dot{\varphi}_k(\tau) = h_l \underline{f}\left(\underline{z}_{lk}, u_{lk_{u,q}}, \underline{\theta}, \tau\right)$$

$$\begin{gathered} l = 1 \dots N_e, \qquad k = 1 \dots K \\ k_{u,q} = 1 \dots K_{u,1}, \qquad q = 1 \dots N_u \\ \underline{\tilde{z}}_{11} = \underline{\tilde{z}}_0 \\ \underline{\tilde{z}}_{l+1k=0} = \underline{\tilde{z}}_{lk=K}, \qquad l = 1 \dots N_e - 1 \\ u_{l+1k_{u,q}=0} = u_{lk_{u,q}=K_{u,q}}, \qquad l = 1 \dots N_e - 1, \qquad K_{u,q} \geq 1 \end{gathered} \qquad (3.12)$$

In order to solve (3.12), $\dot{\varphi}_k(\tau)$ is previously calculated once offline because it only depends on the known selected collocation points τ_k.

3.2 Formulation as a NLP problem

The OED problem has been formulated as a NLP problem in OP2 through total discretization. In this section, a generalized form of NLP is presented and the methods which solve this problem. The interested reader can find a detailed analysis of this topic in (Floudas & Panos 2009) of which the following short introduction is based on. The formulated NLP problem OP2 can be represented by the following generalized formulation with equality and inequality constraints.

$$\begin{aligned} & \min \Phi(\bar{x}) \\ \text{Subject} \quad & 0 = \bar{c}_E(\bar{x}) \\ \text{to:} \quad & 0 \leq \bar{c}_I(\bar{x}) \end{aligned} \qquad (3.13)$$

It is useful to introduce the Lagrangian $\mathcal{L}$ so as to formulate the first and second order conditions of (3.13).

$$\mathcal{L}(\bar{x}, \bar{\lambda}_E, \bar{\lambda}_I) = \Phi(\bar{x}) + \sum_{i=1}^{N_E} \lambda_{E,i} c_{E,i}(\bar{x}) + \sum_{i=1}^{N_I} \lambda_{I,i} c_{I,i}(\bar{x}) \qquad (3.14)$$

Where N_E denotes the number of equality and N_I the number of inequality constraints. With the defined Lagrangian the first-order conditions for (3.13) also known as the KKT conditions are given with the following elements (Luenberger 1984; Nocedal & Wright 2006).

Stationarity condition or also known as balance of forces:

$$\nabla_x \mathcal{L}(\bar{x}^*, \bar{\lambda}_E^*, \bar{\lambda}_I^*) = \nabla \Phi(\bar{x}^*) + \nabla \bar{c}_E(x^*)^T \bar{\lambda}_E^* + \nabla \bar{c}_I(x^*)^T \bar{\lambda}_I^* = 0 \qquad (3.15)$$

Feasibility condition:

$$\begin{aligned} \nabla_{\bar{\lambda}_E} \mathcal{L}(\bar{x}^*, \bar{\lambda}_E^*, \bar{\lambda}_I^*) &= \bar{c}_E(\bar{x}^*) = 0 \\ \nabla_{\bar{\lambda}_I} \mathcal{L}(\bar{x}^*, \bar{\lambda}_E^*, \bar{\lambda}_I^*) &= \bar{c}_I(\bar{x}^*) \leq 0 \end{aligned} \qquad (3.16)$$

Complimentary condition:

$$\begin{aligned} \bar{\lambda}_I^* &\geq 0 \\ c_{I,i}(x^*)^T \bar{\lambda}_I^* &= 0 \end{aligned} \qquad (3.17)$$

The meaning of (3.17) is that either a inequality constraint $c_{I,i}(x)$ is inactive, then the corresponding $\lambda_{I,i}$ must be zero, or active, then by convention, the sign of the corresponding multipliers must be positive. The λs are addressed here as the KKT multipliers. If there are no inequality constraints they are also called as Lagrange-multipliers. Supposed that $\bar{x}^*$ is a local minimum of $\Phi(\bar{x})$ and the LICQ[2] holds. Then, the second order necessary conditions are:

$$\bar{w}^T \nabla_{\bar{x}\bar{x}} \mathcal{L}\left(\bar{x}^*, \bar{\lambda}_E^*, \bar{\lambda}_I^*\right) \bar{w} \geq 0 \qquad \forall \bar{w}, \quad \bar{w} \neq 0 \wedge \nabla c_{E,i}(\bar{x}^*)^T \bar{w} = 0 \qquad (3.18)$$

$$\nabla c_{I,i}(\bar{x}^*)^T w = 0, \quad i \in \{i | c_{I,i}(\bar{x}^*) = 0, \quad \lambda_{I,i}^* > 0\}$$
$$\nabla c_{I,j}(\bar{x}^*)^T w \leq 0, \quad j \in \{j | c_{I,j}(\bar{x}^*) = 0, \quad \lambda_{I,i}^* = 0\} \qquad (3.19)$$

where $\bar{w}$ is a nonzero direction. The same as the second order necessary conditions is required for the sufficient conditions except that (3.18) has to strictly positive definite.

3.3 Solutions to NLP problems

Solutions strategies for NLP problems have been described and explained in detail in (Floudas & Gounaris 2008; Floudas & Panos 2009). There are basically three well studied fundamental approaches to solve the constrained NLP problem represented by OP2, which are the Successive Quadratic Programming, the Interior Point and the Nested strategies (Nocedal & Wright 2006). The following short listing of these strategies are mainly quoted from (Biegler 2010).

1) Successive Quadratic Programming methods (SQP) – are the most well-known and popular algorithm for solving NLP-problems (Bonnans et al. 2009). They basically apply Newton's method to the presented first-order KKT conditions. The assumption is that the active set is known in advance which specifies the "active" constrains with $\lambda_{I,i}^* > 0$ in (3.19). Since the SQP methods are based on Newton's method, they also inherit Newton's method's fast convergence properties. On the other hand, a big disadvantage arises when dealing with large problems since the full Jacobian has to be calculated in each Newton step. If there are few degrees of freedom, then the reduced space SQP (rSQP) method

[2] LICQ: Linear independence constraint qualification – the gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at $\bar{x}^*$.

can be used to overcome this problem since this it takes advantages of sparsity of the gradient. Another main disadvantage is related to poor starting points as known general problem from Newton's method. Some strategies have been introduced so as to ensure stepping towards the optimum based on detecting a feasible step size by using exact penalty functions and on trusted region adaptations. Some of the well-known solvers are presented below:

- SNOPT []
- MUSCOD []
- fmincon []
- …

2) Interior Point methods (IP) – are also known as barrier methods since they are generally using logarithmic penalty/barrier functions (Wright 1987). The advantages are faster convergence for large system with sparse structure and better theoretical properties (global and super-linear convergence). On the other hand new challenges arise with the introduction of logarithmic barrier functions. In particular if the variable becomes very small then its logarithm becomes very large, so that additional strategies have to be involved so as to keep the optimization steps within the "interior" coordinates. Another critical point is the LICQ, which has to be hold during the optimization steps. Some of the well-known algorithms are:

   - IPOPT []
   - KNITRO []
   - LOQO []
   - …

3) Nested methods – are especially developed for dealing with NLP problems where the solver has to remain feasible during the optimization steps. So far, the presented methods basically consider solving the KKT conditions in a simultaneous way but linearization from these conditions at infeasible points can lead to bad search directions and constraint multipliers, and thus, causing the solver to fail. The nested approaches aim at decomposing the main NLP to nested sub-problems and solving them separately, in particular the optimization variables are divided to non-basic, basic and super basic ones. Well investigated algorithms following this philosophy are the Generalized Reduced Gradient and

the Gradient Projection method. Some of the well-known solver using Nested methods are:

- CONOPT []
- MINOS []
- LANCELOT []
- …

3.4 Applied solution strategies

In this work, two kinds of program codes have been implemented so as to solve the OED task as formulated in OP2. The first implementation has been carried out in Matlab using Tomlab's optimizer SNOPT which basically uses an active set strategy while performing the line search with an augmented Lagrangian formulation (Gill et al. 2002). The program structure has been developed as follows:

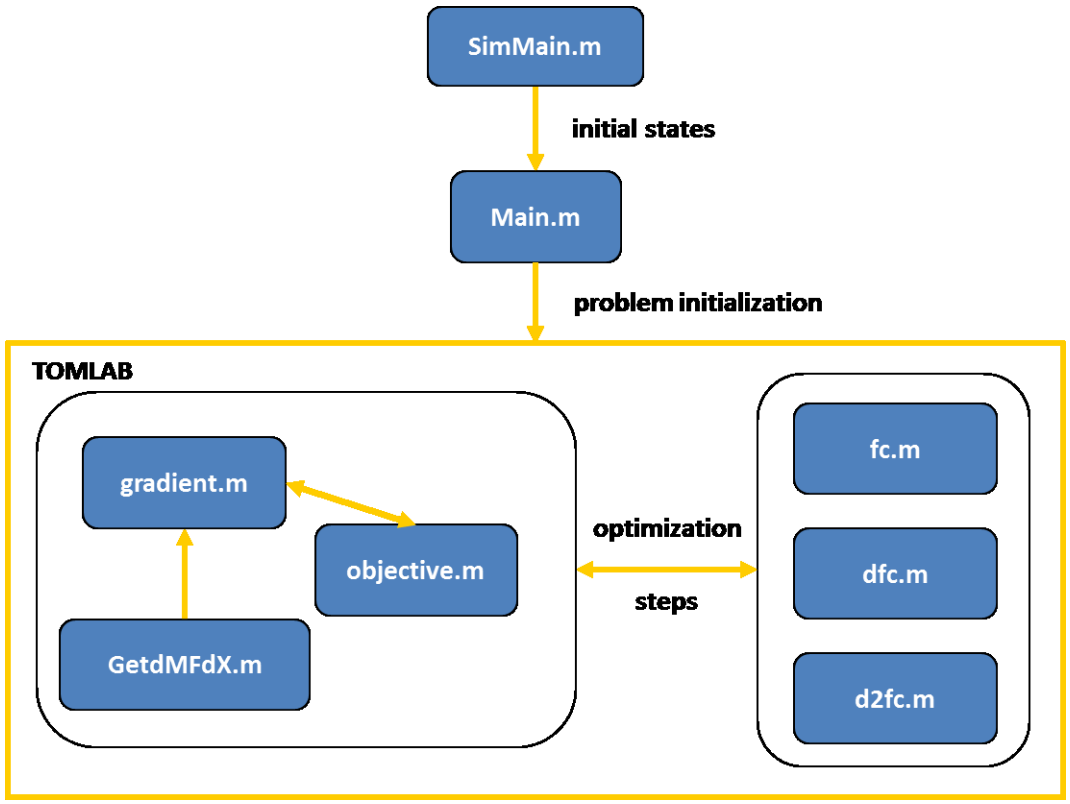


Figure 3.4 Program structure - implementation via SNOPT

The problem formulation OP2is specified in the "Main.m"-file. The initialization vector for the optimizer is given by a previous simulation via the "SimMain.m"-file which solves DAE3 and produces the initial profiles of the states and sensitivities based on the

specified initial conditions with respect to the initial set of dynamic and discrete controls. The main file also specifies the objective criterion which can be chosen as the A-, D- or E- criterion. The optimization structure for Tomlab requires beside the objective function also its first derivative which is calculated via the chain rule as presented in (2.62). The second term of this equation as specified in (2.67) has to be pre-calculated only once and is then stored in the "GetdMFdX.m"-file where it is evaluated during each optimization step. Furthermore, the process constraints and its first and second derivatives as derived in section 2.5 are calculated via the files "fc.m", "dfc.m" and "d2fc.m" respectively.

The second implementation has been carried out in GAMS using the global optimizer BARON with IPOPT as the specified inner NLP solver (Sahinidis 1996). The formulation in GAMS is much closed to OP2 so that it can be treated as a direct transcription (Bruce 2013). The only part which has to be additionally provided is with respect to the discretization orders of the states and the dynamic controls. The only drawback here is that only the A-optimal criterion has been implemented up to now because of the limited possibilities of GAMS to express a matrix, its determinant and eigenvalue respectively.

3.5 Effect of the decision variables to the optimal solution

This section gives a short analysis without prove how the decision variables effects the optimal solution of the OED formulation. In particular, it provides a reference for a qualitative comparison between the sequential and the simultaneous optimization approaches. It is well known from the optimal control theory that for optimality the control profile must result in a bang-bang solution if it linearly enters the Hamilton function (Hermes & LaSalle 1969).

$$\mathrm{H}\big(\mathrm{t},\mathrm{z}(\mathrm{t}),\mathrm{u}(\mathrm{t}),\lambda(\mathrm{t})\big)=\phi\big(\mathrm{t},\mathrm{z}(\mathrm{t}),\mathrm{u}(\mathrm{t})\big)+\lambda(\mathrm{t})\mathrm{f}\big(\mathrm{t},\mathrm{z}(\mathrm{t}),\mathrm{u}(\mathrm{t})\big) \qquad (3.20)$$

With ϕ as the argument of the integral, which has to be maximized, f the right hand sight of the corresponding constraint differential equation and the admissible space of the control as $\mathrm{U_{ad}}=[0,\mathrm{u_{max}}]$. If H is linear in u then (3.20) can be rewritten as

$$\mathrm{H}\big(\mathrm{t},\mathrm{z}(\mathrm{t}),\mathrm{u}(\mathrm{t}),\lambda(\mathrm{t})\big)=\mathrm{g}\big(\mathrm{t},\mathrm{z}(\mathrm{t}),\lambda(\mathrm{t})\big)\mathrm{u}(\mathrm{t})+\mathrm{r}\big(\mathrm{t},\mathrm{z}(\mathrm{t}),\lambda(\mathrm{t})\big) \qquad (3.21)$$

From (3.21) one can easily see that H is a line regarding its argument u with g as the slope. Now, it depends on whether the slope is negative or positive that the optimum is reached at $u = 0$ or $u = u_{max}$, and thus, resulting in a bang-bang solution. A very common case in chemical engineering is that $U_{ad} = [u_{min}, u_{max}]$ and $u_{max} > u_{min} > 0$ due to physical limitations. Then it depends on whether the slope is negative or positive that the optimum is reached at $u = u_{min}$ or $u = u_{max}$. It is analogous regarding the Lagrangian formalism when solving OP2.

The decision variables of OP2 are the sampling decisions $\omega_{i,lk}$ for each state and at each collocation point, the initial condition $z_{i,0}$ of each state, the decision intervals length h_l and the control decisions u_{c,lk_c} for each control at each corresponding collocation point. It is easy to show that $z_{i,0}$ linearly enters the Lagrangian (3.14) since the constraints for the initial conditions are very simple linear equality constraints as seen in (3.22).

$$c_{z_{i,0}} = z_{i,0} - z_{i,11} = 0 \qquad (3.22)$$

It is similar for the decision intervals length h_l as presented in (3.12).

$$\underline{c}_{h_l} = \underline{f}\left(\underline{z}_{lk}, u_{lk_{u,q}}, \underline{\theta}, \tau\right) h_l - \sum_{k=0}^{K} B\left(\underline{z}_{lk}, u_{lk_u}, \bar{\theta}\right) \underline{z}_{lk} \dot{\varphi}_k(\tau) \qquad (3.23)$$

$$l = 1 \dots N_e, \qquad k = 1 \dots K$$

It is more complex to show that the optimal solution for the sampling decisions also results in the bang-bang type since they quadratically enter the objective function $\Phi(M_F)$. Without loss of generality it is only illustrated here for the A- and D-optimal criterion since the objective function can be calculated explicitly for these two criteria. Assume that the process includes two states, two parameters and the collocation scheme uses one discretization element with the polynomial order $K = 2$. Furthermore for the sake of simplicity the measurement covariance matrix is set to the identity matrix so that the FIM can be calculated as

$$M_F = \left(S_{sp}\right)^T S_{sp} = \begin{bmatrix} m_{F11} & m_{F12} \\ m_{F21} & m_{F22} \end{bmatrix} \qquad (3.24)$$

$$S_{sp} = \begin{bmatrix} \omega_{1,11}s_{11,11} & \omega_{1,11}s_{12,11} \\ \omega_{2,12}s_{21,11} & \omega_{2,12}s_{22,11} \\ \omega_{1,11}s_{11,12} & \omega_{1,11}s_{12,12} \\ \omega_{2,12}s_{21,12} & \omega_{2,12}s_{22,12} \end{bmatrix}$$

$$m_{F11} = s_{11,11}^2\omega_{1,11}^2 + s_{11,12}^2\omega_{1,12}^2 + s_{21,11}^2\omega_{2,11}^2 + s_{21,12}^2\omega_{2,12}^2$$

$$m_{F12} = m_{F21} = s_{11,11}s_{12,11}\omega_{1,11}^2 + s_{11,12}s_{12,12}\omega_{1,12}^2 + s_{21,11}s_{22,11}w\omega_{2,12}^2 + s_{21,12}s_{22,12}\omega_{2,12}^2$$

$$m_{F22} = s_{12,11}^2\omega_{1,11}^2 + s_{12,12}^2\omega_{1,12}^2 + s_{22,11}^2\omega_{2,11}^2 + s_{22,12}^2\omega_{2,12}^2$$

Thus, the A-optimal criterion results with respect to $\omega_{1,11}$ in

$$\Phi_A(M_F) = \mathrm{tr}(M_F) = \left(s_{11,11}^2 + s_{12,11}^2\right)\omega_{1,11}^2 + \left(\left(s_{11,12}^2 + s_{12,12}^2\right)\omega_{1,12}^2 + \left(s_{21,11}^2 + s_{22,11}^2\right)\omega_{2,11}^2 + \left(s_{21,12}^2 + s_{22,12}^2\right)\omega_{2,12}^2\right) \quad (3.25)$$

and it is for the D-optimal criterion

$$\Phi_D(M_F) = \det(M_F)$$

$$\begin{aligned}
= (&s_{11,11}^2 s_{12,12}^2 \omega_{1,12}^2 + s_{11,11}^2 s_{22,11}^2 \omega_{2,11}^2 + s_{11,11}^2 s_{22,12}^2 \omega_{2,12}^2 \\
&+ s_{11,11}^2 s_{22,12}^2 \omega_{2,12}^2 + s_{11,11}^2 s_{22,12}^2 \omega_{2,12}^2 + s_{11,12}^2 s_{12,11}^2 \omega_{1,12}^2 \\
&+ s_{12,11}^2 s_{21,11}^2 \omega_{2,11}^2 + s_{12,11}^2 s_{21,12}^2 \omega_{2,12}^2 \\
&- 2 s_{11,11} s_{11,12} s_{12,11} s_{12,12} \omega_{1,12}^2 \\
&- 2 s_{11,11} s_{12,11} s_{21,11} s_{22,11} \omega_{2,11}^2 \\
&- 2 s_{11,11} s_{12,11} s_{21,12} s_{22,12} \omega_{2,12}^2) \omega_{1,11}^2 \\
&+ (s_{11,12}^2 s_{22,11}^2 \omega_{1,12}^2 \omega_{2,11}^2 + s_{11,12}^2 s_{22,12}^2 \omega_{1,12}^2 \omega_{2,12}^2 \\
&+ s_{12,12}^2 s_{21,11}^2 \omega_{1,12}^2 \omega_{2,11}^2 + s_{12,12}^2 s_{21,12}^2 \omega_{1,12}^2 \omega_{2,12}^2 \\
&+ s_{21,11}^2 s_{22,12}^2 \omega_{1,12}^2 \omega_{2,12}^2 + s_{21,12}^2 s_{22,11}^2 \omega_{2,11}^2 \omega_{2,12}^2 \\
&- 2 s_{11,12} s_{12,12} s_{21,11} s_{22,11} \omega_{1,12}^2 \omega_{2,11}^2 \\
&- 2 s_{11,12} s_{12,12} s_{21,12} s_{22,12} \omega_{1,12}^2 \omega_{2,12}^2 \\
&- 2 s_{21,11} s_{21,12} s_{22,11} s_{22,12} \omega_{2,11}^2 \omega_{2,12}^2)
\end{aligned} \qquad (3.26)$$

From this explicit calculation it can be found that $\omega_{1,11}$ quadratically enter the objective function hence the Lagrangian is also quadratic in $\omega_{1,11}$. Analogously, it can be shown for all other $\omega_{i,lk}$. Secondly, $\omega_{i,lk} \in [0,1]$ making them as a quadratic function strict monotonous. For a better understanding, one can substitute $\omega_{i,lk}^2$ by $\omega_{i,lk}^*$. Since the Lagrangian is linear in $\omega_{i,lk}^*$ and because the strict monotonicity of $\omega_{i,lk}^2$ in $[0,1]$ the optimal solution results in a bang-bang solution with $\omega_{i,lk}^* = 0$ or $\omega_{i,lk}^* = 1$, and hence, also providing $\omega_{i,lk} = 0$ or $\omega_{i,lk} = 1$.

It is not trivial if the objective function is formulated as a function of the covariance matrix C_θ. The A-optimal criterion with respect to C_θ is obtained as

$$\begin{aligned}
\Phi_A(C_\theta) &= \operatorname{tr}(M_F^{-1}) \\
&= \frac{(m_1 \omega_{1,11}^2 + n_1)}{(m_2 \omega_{1,11}^2 + n_2)}
\end{aligned} \qquad (3.27)$$

with

$$
\begin{aligned}
m_1 &= \left(s_{11,11}^2 + s_{12,11}^2\right) \\
n_1 &= \left(s_{11,12}^2 + s_{12,12}^2\right)\omega_{1,12}^2 + \left(s_{21,11}^2 + s_{22,11}^2\right)\omega_{2,11}^2 \\
&\quad + \left(s_{21,12}^2 + s_{2212}^2\right)\omega_{2,12}^2
\end{aligned}
\tag{3.28}
$$

and

$$
\begin{aligned}
m_2 = \big(&s_{11,11}^2 s_{12,12}^2 \omega_{1,12}^2 - 2s_{11,11}s_{11,12}s_{12,11}s_{12,12}\omega_{1,12}^2 + s_{11,11}^2 s_{22,11}^2 \omega_{2,11}^2 \\
&+ s_{11,11}^2 s_{2212}^2 \omega_{2,12}^2 + s_{11,12}^2 s_{12,11}^2 \omega_{1,12}^2 + s_{12,11}^2 s_{21,11}^2 \omega_{2,11}^2 \\
&+ s_{12,11}^2 s_{21,12}^2 \omega_{2,12}^2 - 2s_{11,11}s_{11,12}s_{21,11}s_{22,11}\omega_{2,11}^2 \\
&- 2s_{11,11}s_{11,12}s_{21,12}s_{22,12}\omega_{2,12}^2\big) \\
n_2 = \big(&s_{11,12}^2 s_{22,11}^2 \omega_{1,12}^2 \omega_{2,11}^2 + s_{11,12}^2 s_{2212}^2 \omega_{1,12}^2 \omega_{2,12}^2 \\
&+ s_{12,12}^2 s_{21,11}^2 \omega_{1,12}^2 \omega_{2,11}^2 + s_{12,12}^2 s_{21,12}^2 \omega_{1,12}^2 \omega_{2,12}^2 \\
&+ s_{21,11}^2 s_{22,12}^2 \omega_{2,11}^2 \omega_{2,12}^2 + s_{21,12}^2 s_{22,11}^2 \omega_{2,11}^2 \omega_{2,12}^2 \\
&- 2s_{11,12}s_{12,12}s_{21,11}s_{22,11}\omega_{1,12}^2 \omega_{2,11}^2 \\
&- 2s_{11,12}s_{12,12}s_{21,12}s_{22,12}\omega_{1,12}^2 \omega_{2,12}^2 \\
&- 2s_{21,11}s_{21,12}s_{22,11}s_{22,12}\omega_{2,11}^2 \omega_{2,12}^2\big)
\end{aligned}
\tag{3.29}
$$

The D-optimal criterion with respect to C_θ is

$$
\begin{aligned}
\Phi_D(C_\theta) &= \mathrm{tr}(M_F^{-1}) \\
&= \frac{1}{\left(m_2\omega_{1,11}^2 + n_2\right)}
\end{aligned}
\tag{3.30}
$$

Now, it depends on the signs and the magnitudes of m_2 and n_2 whether $\Phi_A(C_\theta)$ and $\Phi_D(C_\theta)$ have a pole in [0,1] or not. Hence strict monotonicity cannot be guaranteed as it has been shown for the formulation with the FIM, in particular with $\Phi_A(M_F)$ and $\Phi_D(M_F)$.

If the control decisions u_{q,lk_q} linearly enter the constraints in (2.58) which is a priori determined by the process model, then they also has to results in a bang-bang solution for optimality if no other constraints on the state variables are violated.

3.6 Possibilities of the applied solution approaches

In this chapter, a more detailed investigation than the introduction section is presented regarding the possibilities and advantages of the proposed simultaneous optimization approach to OED. The results are based on the application of the engineering example of a biomass reactor in section 1.3 with the difference that now all four parameters $\theta_{1\ldots4}$ are unknown. Thus the objective function is formulated with respect to all four parameters. The first two subsections take into account the gain from the variation of initial states as well as variation of dynamic control types. Subsection 0 gives a detailed insight to the outcomes regarding the OASE where the implementation with the global optimizer BARON clearly is superior over the implementation with SNOPT. The last part additionally presents optimization results with respect to the D- and E criterion obtained via SNOPT.

3.6.1 Optimization of the initial conditions

The semi-batch biomass reactor is optimized with zero-order control functions. The focus here is on the optimized initial conditions, thus, the total experiment time and the sampling decisions are fixed. The reference criterion is chosen as A-optimal. The initial settings are summarized in Table 9.

Table 9 Initial settings – Variation of control orders

process controls		process variables	
$u_1(t) = \text{const.}$	$0.05\ h^{-1}$	$c_B(t = 0)$	$1\ g/l$
$u_2(t) = \text{const.}$	$0.2\ g/l$	$c_S(t = 0)$	$25\ g/l$
measurement variance		$\sigma_B = 1\ g/l,$	$\sigma_S = 1\ g/l$
experiment duration	$10\ h$		
sampling vector [h]	$[2, 4, 6, 8, 10]$		

The OED formulation OP2 for this problem is implemented via Tomlab's optimizer SNOPT as presented in 3.4. The resulting control trajectories are shown in Figure 3.5 and Figure 3.6 respectively.

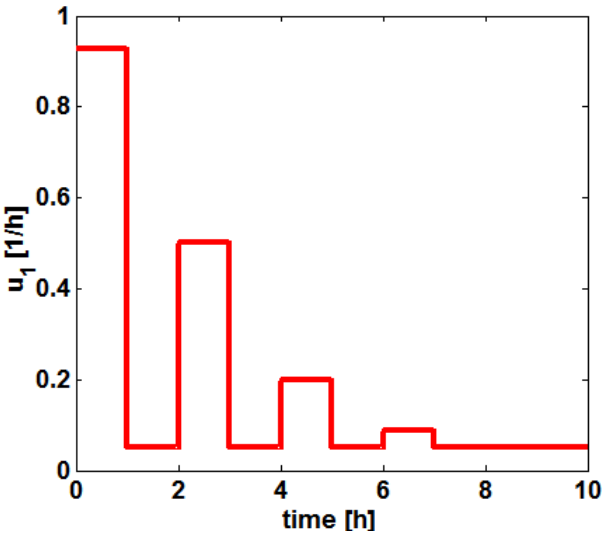


Figure 3.5 Control trajectory of the dilution factor, optimized initial conditions

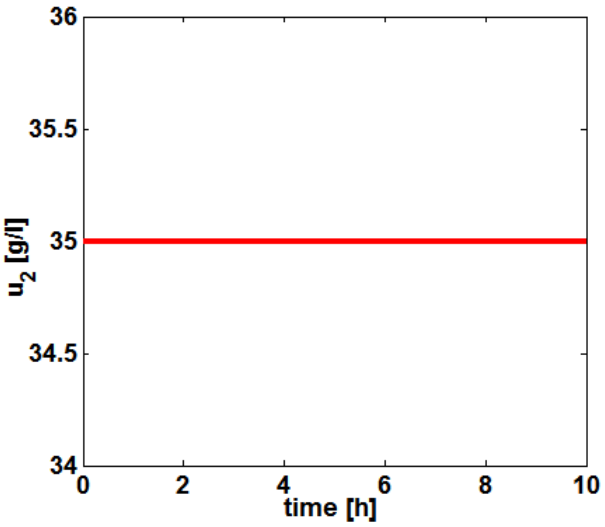


Figure 3.6 Control trajectory of the feed substrate, optimized initial conditions

The resulting profiles of the state variables are shown in Figure 3.7 and Figure 3.8. As discussed in 3.5, the optimized initial conditions are optimal since they both hit their bounds.

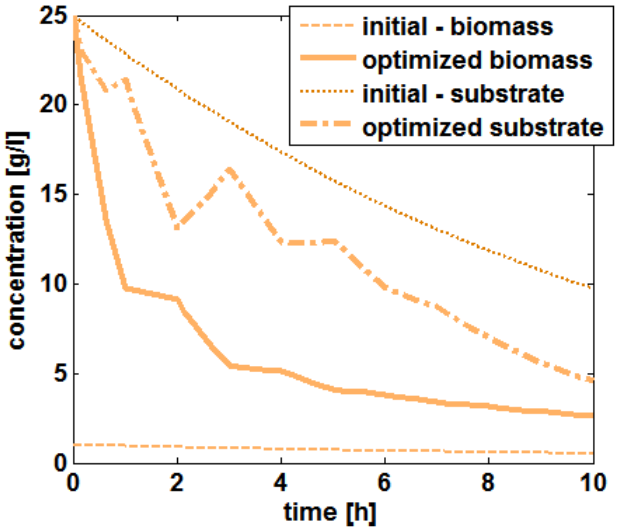


Figure 3.7 Concentration profiles, initial vs. optimized settings

The information content represented by the A-criterion from the initial settings has a nominal value of $1.116 \cdot 10^3$. After applying the OED approach with optimized initial conditions and dynamic control trajectories the A-criterion could be clearly increased as seen in in Table 10.

Table 10 Optimization results – optimized initial conditions

	initial settings	optimized initial conditions
A-criterion	$1.116 \cdot 10^3$	$2.593 \cdot 10^4$

3.6.2 Variation of the dynamic control variables

In practice zero and first-order control types are most relevant since the implementation of control types of higher order are hardly to implement. The previous problem is solved with first-order control for the dilution factor and zero-order control for the feed substrate in the first case and in the second case with first-order for both controls respectively.

In both cases the optimized control trajectories for the dilution factor result in the same profile as seen in Figure 3.8. More important is that even with a first-order control type in the second case, the optimized control trajectory for the feed substrate also hit the upper constraint, thus, is considered as optimal.

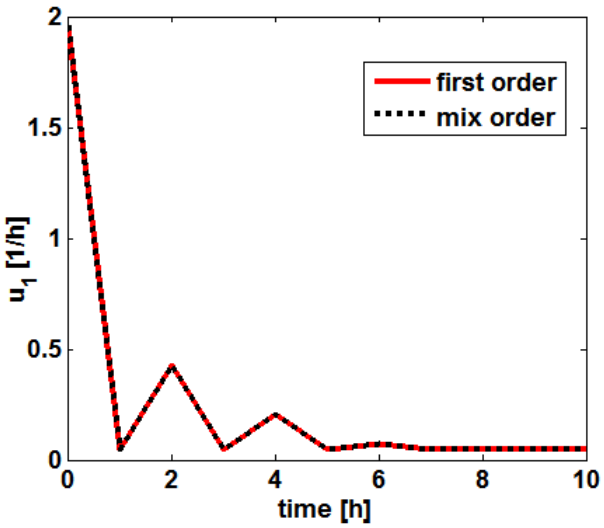


Figure 3.8 Control trajectory of the dilution factor, variation of control orders

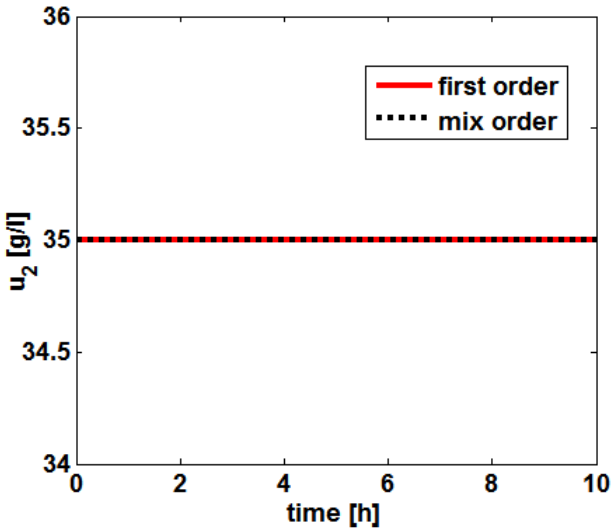


Figure 3.9 Control trajectory of the feed substrate, variation of control orders

The corresponding profiles of the state variables are shown in Figure 3.10. A comparison regarding optimized criterion between both results and the previous case, where only zero-order controls were used, is summarized in Table 11.

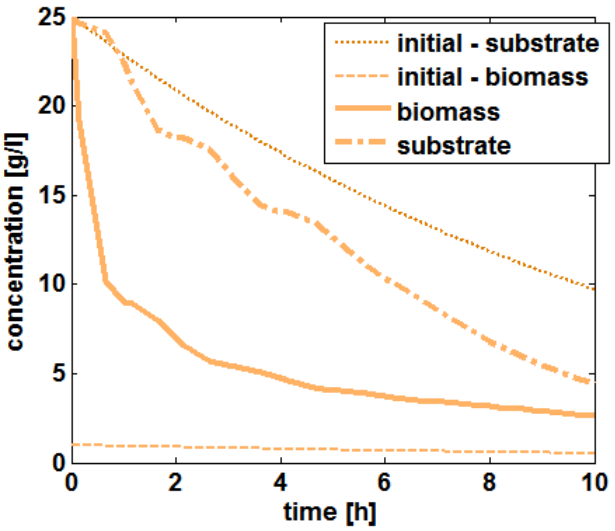


Figure 3.10 Concentration profiles, variation of control orders

Table 11 Optimization results – comparison of different control order

	zero-order	first order	mix-order
A-criterion	$2.593 \cdot 10^4$	$2.563 \cdot 10^4$	$2.563 \cdot 10^4$

This subsection has given a short investigation of the effect of different control orders. In theory, the optimal case is generally obtained with a step-switching structure which results in a bang-bang solution, if the controls linearly enter the process model and there is no resulting violation on the constraints of the state variables. Indeed one can recognize from Figure 3.5 and Figure 3.8 that the optimized control trajectories of the dilution factor with mix- and only first-order controls attempt to reach the solution of the only zero-order case, which gives the best nominal value for the objective criterion. However, step functions cannot be implemented in practice, for example regarding temperature, flow rates and stirrer speed. In these cases, first-order controls are essential and the compromise between optimality and practical implementation has to be taken in to account.

3.6.3 Optimal adaptive sampling Strategy – OASE

The optimization problem of the previous subsections is now extended with more degrees of freedom from the OASE. In particular, the sampling decisions ω and the duration of each decision interval h_l are also optimized. First, the OED formulation of OP2 is solved with the implementation in Tomlab via SNOPT, then with the implementation in GAMS via BARON. The initial conditions, the initial control settings, as well as additional constraints are summarized in Table 12.

Table 12 Initial conditions and initial control settings

process controls		process variables	
$u_1(t) = \text{const.}$	$0.05\ h^{-1}$	$c_B(t = 0)$	$1\ g/l$
$u_2(t) = \text{const.}$	$0.2\ g/l$	$c_S(t = 0)$	$25\ g/l$
u_1 control type		first-order– ramp	
u_2 control type		zero-order - step	
experiment time limit		$t_{end} \leq 20\ h$	
constraints on decision interval		$1h \leq h_l \leq 2h$	
sampling number		5 for each state	
measurement variance		$\sigma_B = 1\ g/l, \quad \sigma_S = 1\ g/l$	
initialization of the sampling vector [h]		$[\ 2, 4, 6, 8, 10\ ]$	
Objective criterion		A-optimal	

Results via SNOPT

The optimized control trajectories are shown in Figure 3.11 and Figure 3.12. They are very close to the ones of the previous section with mix-order controls.

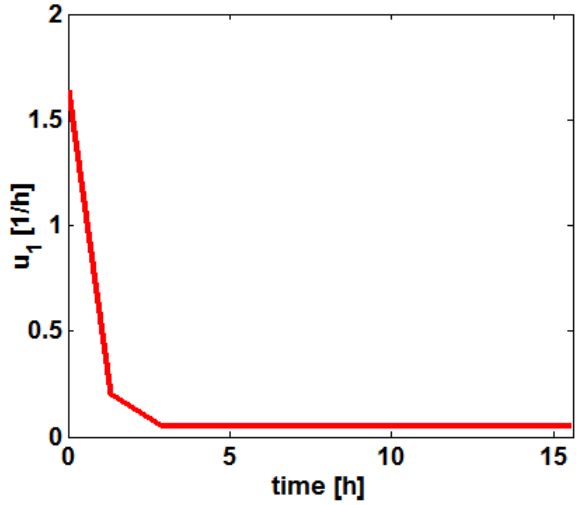


Figure 3.11 Control trajectory of the dilution factor, OASE - SNOPT

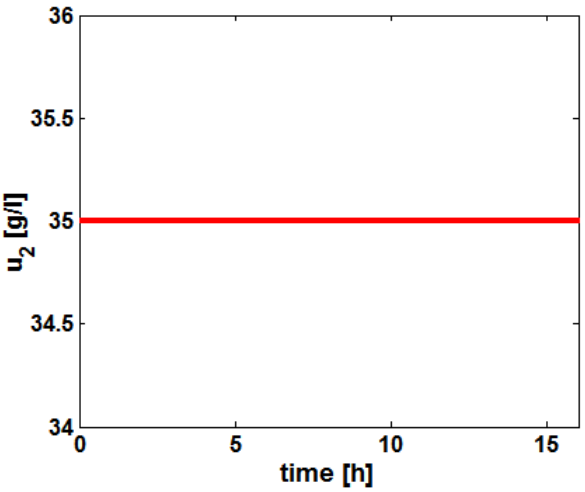


Figure 3.12 Control trajectory of the feed substrate, OASE–SNOPT

The corresponding state variable profiles are shown in Figure 3.13. It can be clearly seen from the comparison with the initial settings that the experiment duration has been increased to 15.6 hours but not reaches the allowed maximum time of 20 hours. Figure 3.14 shows the decision intervals duration, which is clearly not a bang-bang solution as expected from the discussion in Figure 3.5, thus, the obtained solution is not “global” optimal with respect to the decision intervals duration.

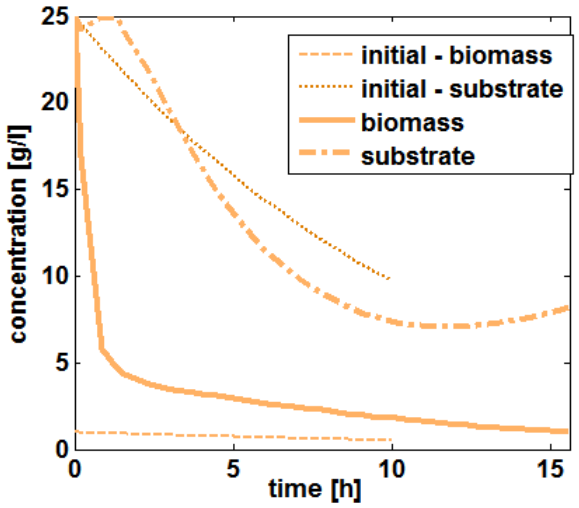


Figure 3.13 Concentration profiles, OASE–SNOPT

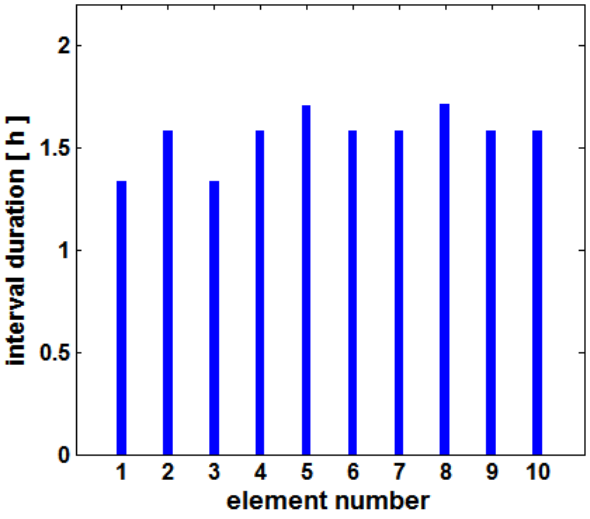


Figure 3.14 Decision intervals, OASE – SNOPT

The OASE for the measured variables is illustrated in Figure 6.1 and Figure 6.2. The optimized criterion via SNOPT is compared with the mix-order approach of the previous subsection. The results show that the nominal value of the A-criterion could be improved with the OASE by a factor of two.

Table 13 Optimization results – OASE via SNOPT

	initial settings	mix-order	SNOPT
A-criterion	$1.116 \cdot 10^3$	$2.563 \cdot 10^4$	$5.013 \cdot 10^4$

Results via BARON

The optimized control trajectories show the same behavior compared to the optimization results via SNOPT as seen in Figure 3.15 and Figure 3.16.

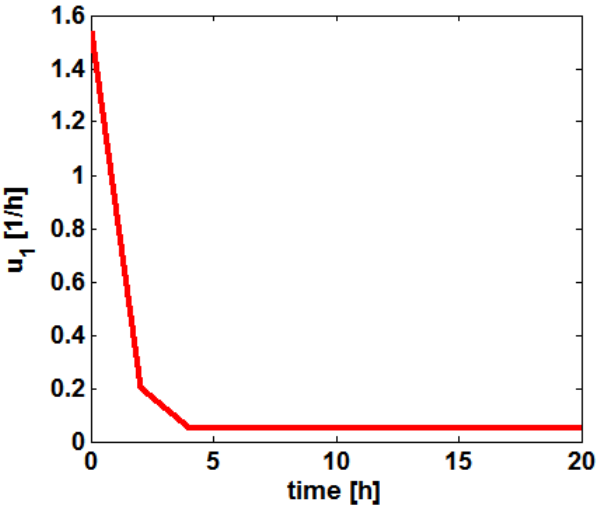


Figure 3.15 Control trajectory of the dilution factor, OASE - BARON

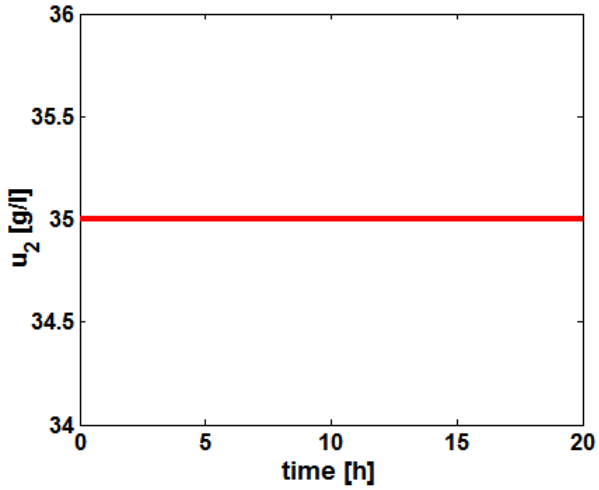


Figure 3.16 Control trajectory of the feed substrate, OASE – BARON

The corresponding profiles of the state variables in Figure 3.17 show more clearly that the trends of the results of both optimization approaches are quite similar. The difference lies in the optimal decision intervals which hit the bounds as seen in Figure 3.18.

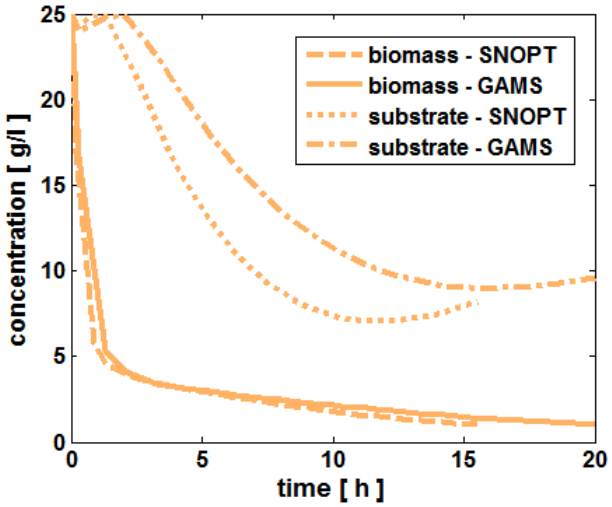


Figure 3.17 Concentration profiles, OASE – BARON

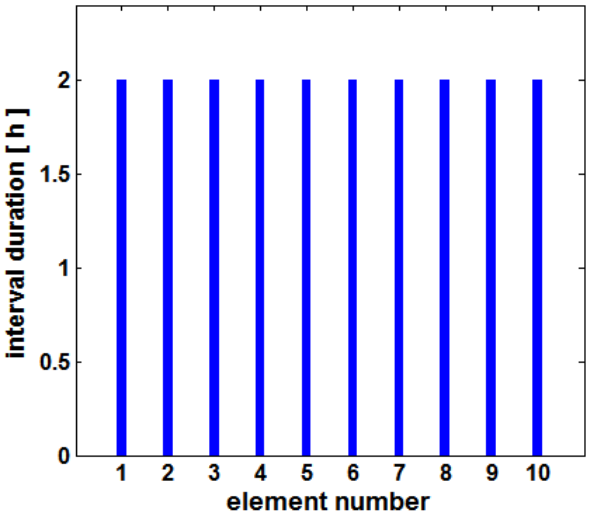


Figure 3.18 Decision intervals, OASE – BARON

The OASE for the measured variables obtained from the optimization with BARON are shown in Figure 6.3 and Figure 6.4. Compared to the results via SNOPT, these results are more reasonable since the measurements are more taken at the end and in the middle of the experiments. Since these are dynamic processes, the largest difference of the model output with two different parameter set is expected at the end of the process or in the middle of the process, where its dynamic has the largest change but not at the beginning of the process as observed in Figure 6.1.

The plausible qualitative improvement by a better solution regarding decision intervals and OASE leads to a further increase of the nominal value of the objective criterion as seen in Table 14.

Table 14 Optimization results – comparison SNOPT vs. BARON

	initial settings	SNOPT	BARON
A-criterion	$1.116 \cdot 10^3$	$5.013 \cdot 10^4$	$1.001 \cdot 10^5$

3.6.4 Different criteria

This section takes the D- and E-optimal criteria into account. Since these two criteria could not be implemented in GAMS, the optimization results are obtained by the implementation in Tomlab via SNOPT.

D-optimal criterion

First, the optimization has been carried out with respect to the D-optimal criterion.

Figure 3.19 and Figure 3.20 show the optimized control trajectories of first-order type for the dilution factor and zero-order type for the feed substrate. The optimality of the control trajectory of the feed substrate is not given anymore.

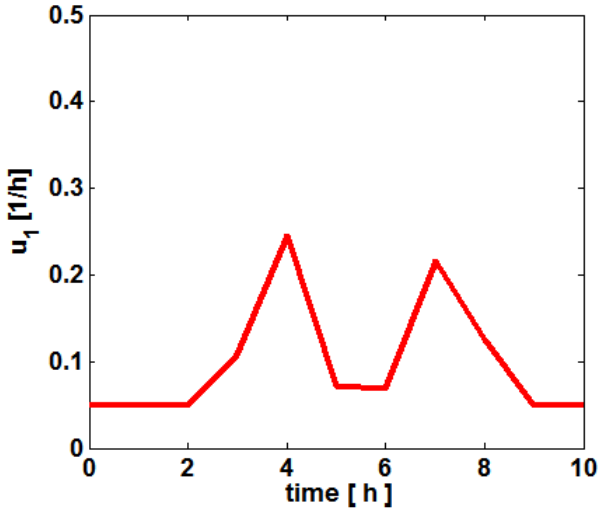


Figure 3.19 Control trajectory of the dilution factor, D-optimal – SNOPT

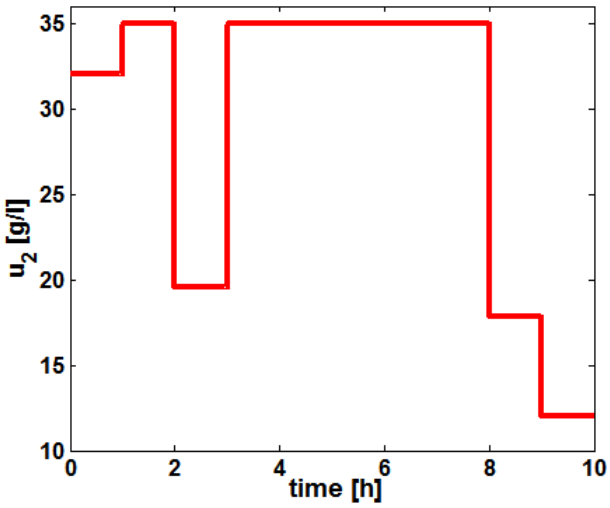


Figure 3.20 Control trajectory of the feed substrate, D-optimal – SNOPT

Both control profiles show a more dynamic behavior, what is typical for OED problems if the D-criterion is used. The optimality with respect to the initial conditions could be only kept for the substrate concentration. The resulting profiles of both states are given in Figure 3.21.

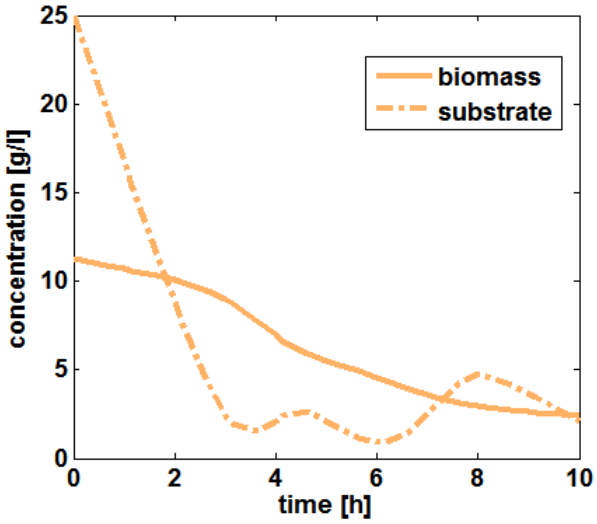


Figure 3.21 Concentration profiles, D-optimal – SNOPT

The optimized decision intervals hit the lower bounds as seen in Figure 6.5. The OASE results are presented in Figure 6.6 and Figure 6.7. Both profiles reasonably put the measurements at dynamic parts as well as at the end part of the experiments. Although optimality could not be obtained for every decision variables the nominal value of the objective criterion has remarkably been increased as seen in Table 15.

Table 15 Optimization results – D-criterion

	initial settings	SNOPT
D-criterion	$1.434 \cdot 10^{-8}$	$1.392 \cdot 10^{6}$

Since the D-criterion is the determinant of the FIM and is obtained basically via a product of the sensitivities a small number of near zero sensitivity values can push the total nominal criterion value to a very small number. Hence an optimized result typically gains large improvements.

E-optimal criterion

In the second part the OED problem is solved with respect to the E-optimal criterion which is preferred if single main principle component of the FIM is distorted (Steinberg & Hunter 1984). However, from a perspective of benchmarking the E-criterion is also useful to evaluate the robustness of the optimizer with respect to handling the system dynamic. Since the minimum eigenvalue of the FIM is maximized and as very well-known from the linear algebra an eigenvalue of a matrix can have a large change by small changes of the matrix an optimizer as SNOPT without global optimum strategy usually gets stuck in local optima and sometime fails to move towards to the global optimum at all.

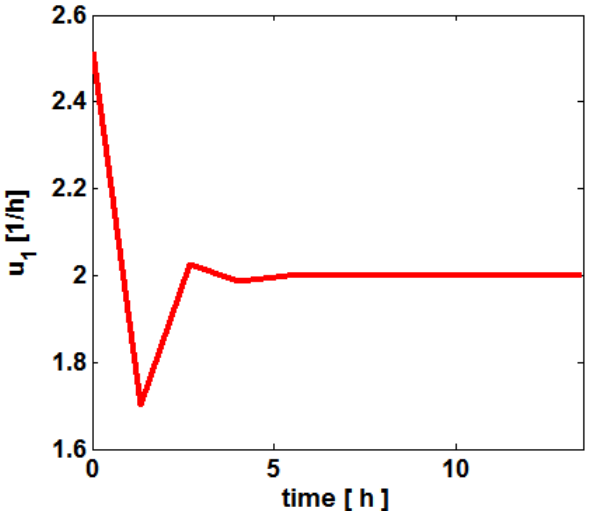


Figure 3.22 Control trajectory of the dilution factor, E-optimal - SNOPT

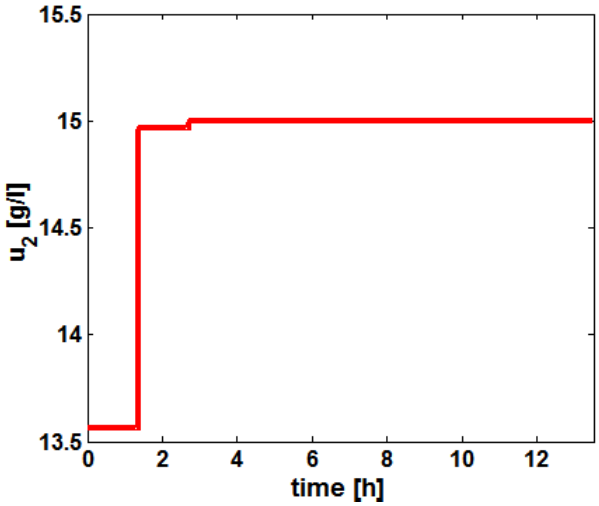


Figure 3.23 Control trajectory of the feed substrate, E-optimal – SNOPT

From the results one can see that all optimality as seen from the results via BARON were lost except for the initial condition of the substrate concentration. The control trajectories are hardly perturbed as presented in Figure 3.22 and Figure 3.23. Hence the system dynamic and the sensitivities cannot fully be addressed leading to an almost stationary process behavior as seen from the concentration profiles in Figure 3.24.

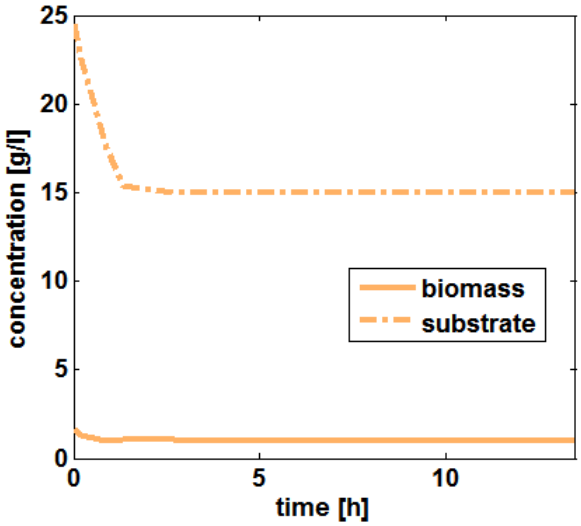


Figure 3.24 Concentration profiles, E-optimal – SNOPT

Furthermore the results with respect to the decision intervals (Figure 6.8) and the OASE (Figure 6.9 and Figure 6.10) also reveals that the optimizer could not proceed many optimization steps since the corresponding decision variables already stuck at the

initialization values. As a result there is almost no information gain as reflected by the nominal criterion value in Table 16.

Table 16 Optimization results – E-criterion

	initial settings	SNOPT
E-criterion	$3.728 \cdot 10^{-11}$	$1.667 \cdot 10^{-8}$

3.7 Summary of the simultaneous optimization approach to OED

This chapter presented the theoretical derivation and the application of the simultaneous optimization approach to OED. It has been shown that the basic concept of total discretization which converts the related optimal control problem into a NLP problem opens many new possibilities in terms of a well formulated optimal experimental design task. The advantages over the conventional sequential approach are the saving of the entire parameter sensitivity differential equation system of higher orders and the sensitivity differential equation system with respect to the initial states. Since the corresponding decision variables explicitly appear in the optimization formulation OP2 their derivatives can be directly obtained through differentiation. Furthermore the handling of flexible order of the control variables is superior to the sequential approach by applying the same discretization scheme as the state variables. Hence additional differential equations and parameterization schemes as known from the implementation with the sequential approach can be avoided.

The effect of the decision variables has been theoretically derived and successively discussed by the application examples. The most important result of this investigation is that the sampling decision variables ω quadratically enter the Lagrangian but are strict monotonous within their bounds. Hence a feasible solution which strictly put the sampling decisions to zero or one can be guaranteed for the formulation of OP2 with respect to the FIM. In contrast, it has been shown that the sampling decisions enter the Lagrangian highly nonlinear if OP2 is formulated with respect to the covariance matrix C_θ. In particular, a strict monotonicity cannot be guaranteed for the whole domain of the sensitivities since poles can arise. Hence the formulation with the FIM is superior and clearly recommended.

The NLP problem in OP2 has been solved with Tomlab's SNOPT which uses the reduced successive quadratic programming method (rSQP) via an implementation in Matlab and with BARON's IPOPT which uses the interior point method via an implementation in GAMS. The best results could be obtained with the implementation in GAMS with BARON, however only the A-criterion could be formulated in this programming language yet. Since BARON is a global optimizer it clearly has the advantage in finding solutions which fulfill the optimal condition always resulting in bang-bang solutions for decision variables which linearly enter the system equations. On the other hand it is possible to solve OP2 with the A-, D- and E-criterion via Tomlab but although optimal solutions cannot be guaranteed for the more dynamic D- and E-criterion.

The advantages of the simultaneous approach to OED over the state of the art sequential approaches are summarized in Table 17. As derived in OP1 and OP2, the implementation of dynamic controls of higher orders and of optimized initial conditions is straight forward in the simultaneous formulation whereas additional equation systems and integration effort are necessary[*1] for the sequential approaches.

Table 17 Advantages of the simultaneous optimization approach to OED

	Single Shooting optimization	Multiple Shooting optimization	Simultaneous optimization
• dynamic controls	✓[*1]	✓[*1]	✓
• optimized initial conditions	✓[*1]	✓[*1]	✓
• constraints w.r.t controls and states	(✓)[*2]	(✓)[*2]	✓
• handling unstable systems	/	(✓)[*2]	✓
• optimal adaptive sampling strategy - OASE	(✓)[*3]	(✓)[*3]	✓

As also shown in OP2 and discussed by the results of section 4.1, the advantages of the simultaneous approach are clearly superior in case of constraints with respect to the state variables. Since it is possible to direct put the crucial process limitation directly as path constraints the simultaneous approach can overcome problems with unstable systems reaching optimal results whereas the sequential approaches generally fail in even moving further than the initialization region [*2]. A reformulation of the optimization problem is also necessary for an implementation of the OASE with sequential approaches. Analogously to the first point, additional equation systems and integration effort are also necessary in this case, whereas it can be directly differentiated with respect to all corresponding optimization variables with the simultaneous approach, hence additional integration effort can be avoided[*3].

4 Application examples

4.1 A theoretical instable CSTR application

The first case study is a continuous stirred-tank reactor, which was firstly treated in (Klatt & Engell 1993) and later reviewed by other authors (Kirches et al. 2012) concerning nonlinear model predictive control. The settings by (Kirches et al. 2012) are modified for the purpose of OED and illustrated in Figure 4.1. The aim of this study is to show the effect of system instabilities on the outcome of the OED under the presence of process constraints.

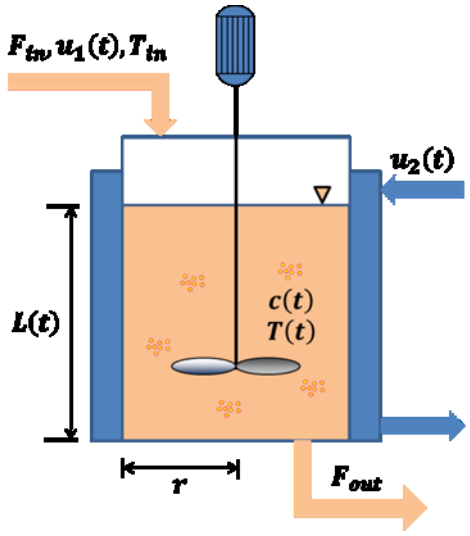


Figure 4.1 CSTR - model

The model equations are formulated as follows:

$$\frac{dc(t)}{dt} = \frac{F_{in}u_1(t) - F_{out}c(t)}{A_R L} - k_{r,0}e^{-\frac{E_a}{RT(t)}}c(t) \qquad (4.1)$$

$$\frac{dT(t)}{dt} = \frac{F_{in}T_{in} - F_{out}T(t)}{A_R L} - \frac{\Delta H_r}{\rho C_p}k_{r,0}e^{-\frac{E_a}{RT(t)}}c(t) + \frac{2U}{r\rho C_p}\left(u_2(t) - T(t)\right) \qquad (4.2)$$

The reactor concentration c and the reactor temperature T are the system state variables. For the target of experimental design the parameter sensitivities of the reaction frequency factor $k_{r,0}$ and the heat transfer coefficient U should be improved for a subsequent parameter estimation task. The control variables of the systems are the inlet

concentration c_{in} and the temperature ofthe cooling water T_{cool}. All the other variables are process constants (see Table 32).

$$u_1(t) = c_{in}(t)$$
$$u_2(t) = T_{cool}(t) \qquad (4.3)$$

For the sake of simplicity, it is assumed that the level is perfectly controlled thus $F_{in}(t) = F_{out}(t)$. The process constraints on state and control variables are:

$$0.8\ \text{mol/l} \leq c(t) \leq 1\ \text{mol/l}$$
$$298\ \text{K} \leq T(t) \leq 333\ \text{K}$$
$$0.8\ \text{mol/l} \leq u_1(t) \leq 1\ \text{mol/l} \qquad (4.4)$$
$$288\ \text{K} \leq u_2(t) \leq 353\ \text{K}$$

The process has a highly nonlinear and unstable characteristic as seen in the simulation results of the real process (see section 6.2). The setting of the discrete control levels and constant initial conditions for an exemplary factorial design approach is given in Table 18.

Table 18 Simulation settings

process controls		process variables	
u_1 [mol/l]	[0.80, 0.85, 0.90, 0.95, 1.00]	c [mol/l]	0.877
u_2 [K]	[288, 293, 298,303, 313, 318, 323,328, 338, 343, 348,353]	T [K]	323
experiment duration [min]	20		

Although the initial conditions have been set constant, the instability of the process is highly significant and present in the whole design variable space. A factorial design approach based on the settings of Table 18 would lead to 59 inadmissible experiments of totally 70, equally to only 16% yield just because the lack of the conventional approach to take process constraints into account. Consequently, the OED approach is used to address this problem. Here, the focus is the comparison between the sequential

and the simultaneous optimization approach. The initialization of the OED task for both cases is summarized in Table 19.

Table 19 Initial conditions and initial control settings

process controls		process variables	
u_1 [mol/l]	0.9	c[mol/l]	0.877
u_2 [K]	300	T[K]	323
experiment time limit		20 min	
sampling number		5 for each state	
measurement variance		$\sigma_c = 1$ mol/l, $\sigma_T = 1$ K	
initialization of the sampling vector [min]		[4, 8, 12, 16, 20]	
Objective criterion		A-optimal	

The simulation results of the state variables with the initialization settings are shown in Figure 4.2 providing an admissible starting operating point.

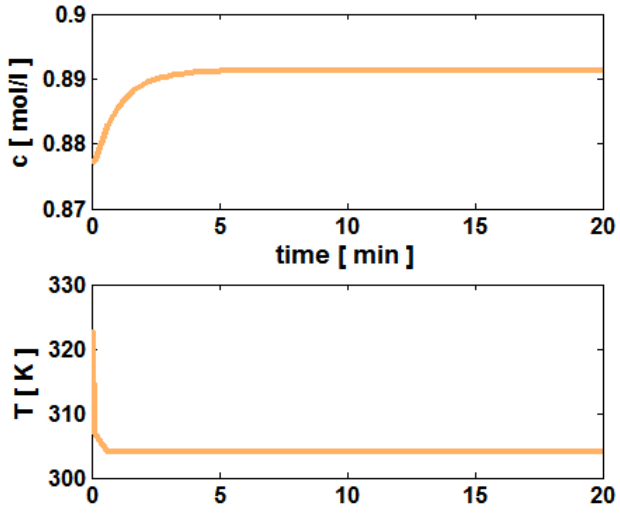


Figure 4.2 CSTR Simulation result - process model with initial settings

Sequential optimization approach

The experimental design package from gPROMS was used as reference. Since the sequential optimization approach is generally not suitable for unstable processes, the integrated optimizer totally fails after exceeding the iteration steps limit and could not find an admissible solution.

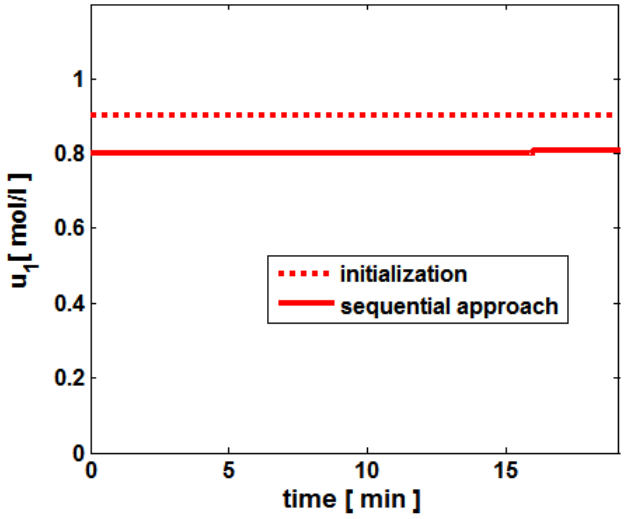


Figure 4.3 Control trajectory of the inlet concentration, sequential approach

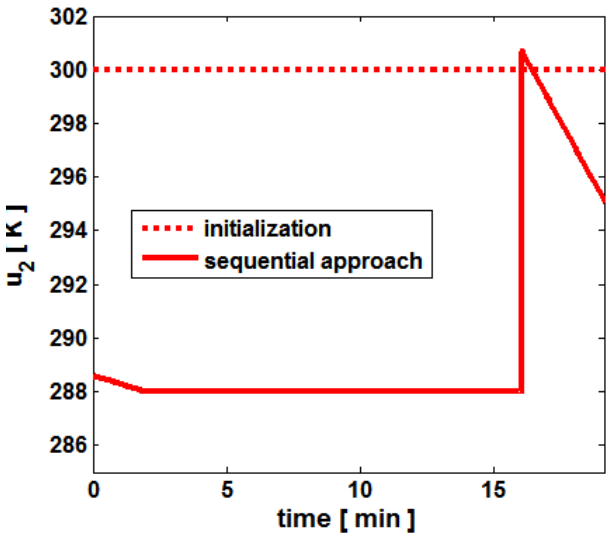


Figure 4.4 Control trajectory of the cooling temperature, sequential approach

The initial settings for both controls have not been varied much as seen in Figure 4.3 and Figure 4.4 since the process constraints on the state variables were aggressively violated for the entire experiment time as also seen in Figure 4.5 and Figure 4.6.

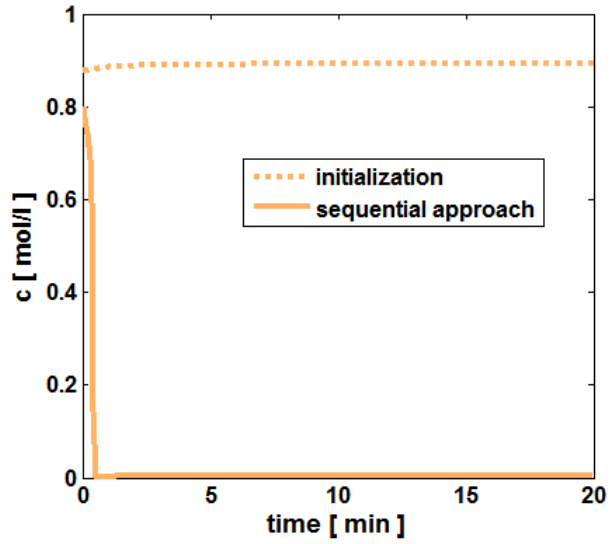


Figure 4.5 Reactor concentration profile, sequential approach

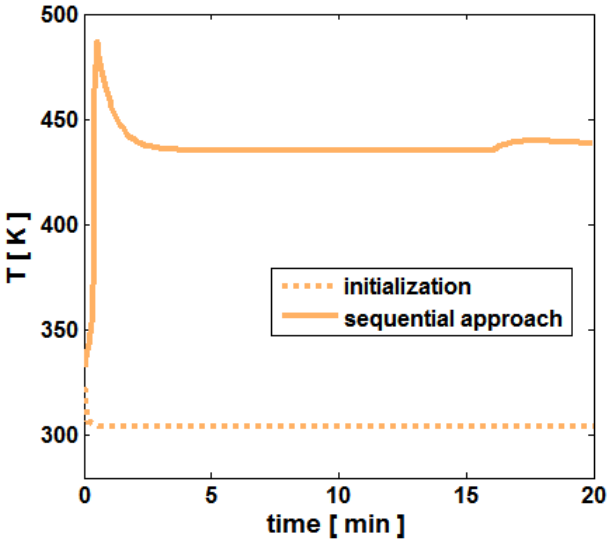


Figure 4.6 Reactor temperature profile, sequential approach

Furthermore the optimizer cannot lead the process back to stability after entering an unstable region. The corresponding decision interval lengths and the OASE are shown in Figure 6.11 and Figure 6.12, which has just a marginal meaning since the optimization result violates the constraints.

Simultaneous optimization approach

The optimized control trajectory for the inlet concentration u_1 is kept all the time at its upper bound representing a bang-bang solution as seen in Figure 4.7. Since u_1 linearly enters the model equation (4.1) this result is optimal as discussed in section 3.5.

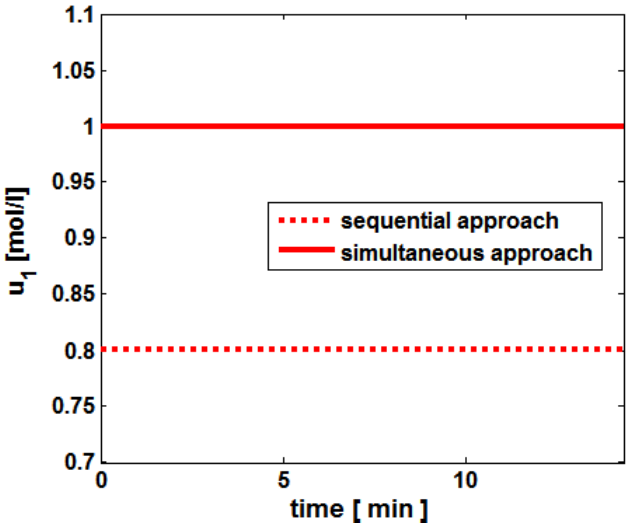


Figure 4.7 Control trajectory of the inlet concentration, simultaneous approach

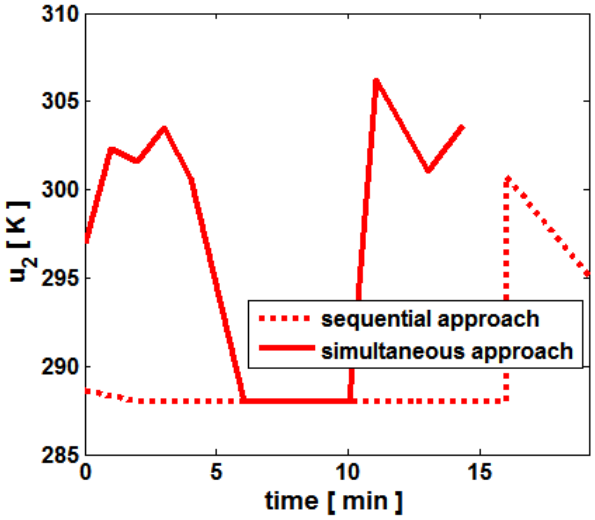


Figure 4.8 Control trajectory of the cooling temperature, simultaneous approach

In contrast to the sequential approach the simultaneous approach results in a reasonable solution where the state variables are kept within the process constraints through the entire experiment as seen in Figure 4.9 and Figure 4.10.

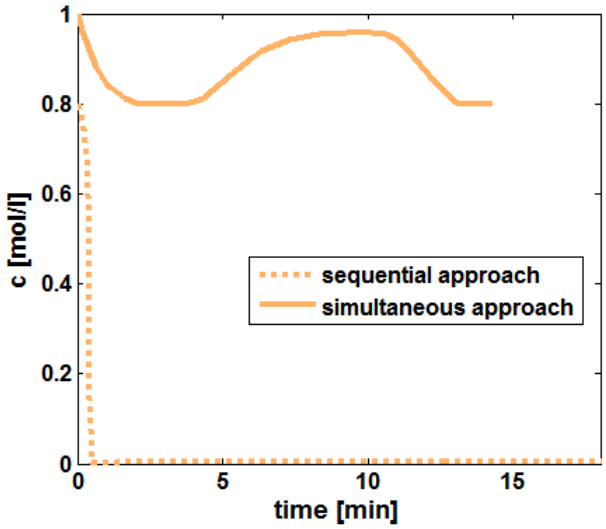


Figure 4.9 Reactor concentration profile, simultaneous approach

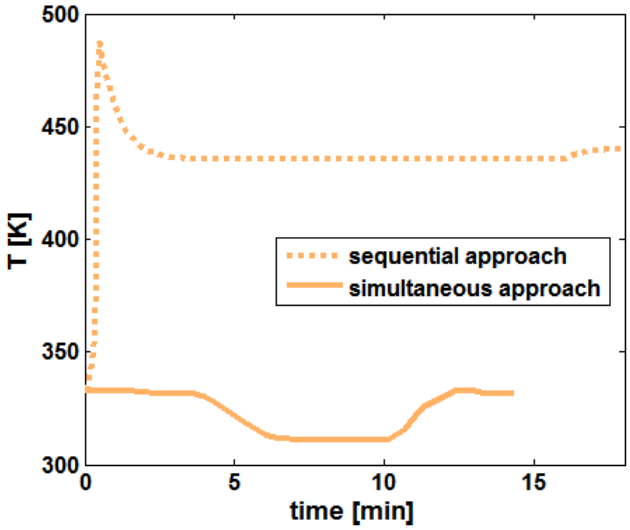


Figure 4.10 Reactor temperature profile, simultaneous approach

The corresponding decision interval lengths and the OASE are shown in Figure 6.13 and Figure 6.14 respectively. The sampling decisions show a proper strategy for a dynamic process where information is generally taken during and at the end part of the experiments.

4.2 Hydroformylation application

The second case study corresponds to a hydroformylation process of 1-dodecene, a homogeneous catalyzed reaction via transition metals (Haumann et al. 2002; Leeuwen & Claver 2008; Koeken et al. 2011). The goal of this engineering example is to show an entire parameter estimation procedure of a complex reaction kinetic supported by OED. The main goals of this process focus on an economical catalyst recycling and a high selectivity regarding the linear aldehyde product n-tridecanal (Kraume 2013). For this purpose, the hydroformylation process was successfully investigated[3] in a DMF-decane thermomorphic solvent (TMS) system with a Biphephos-modified rhodium catalyst (Brunsch & Behr 2013) and in a Marlophen surfactant system with a Sulfoxantphos-modified rhodium catalyst (Rost et al. 2013).

It is well known that side reactions like isomerization and hydrogenation of 1-dodecene and its isomers occur during the process. Furthermore, side- or subsequent reactions leading to aldols, alcohols, and acids may occur. However, since the latter side reactions have not been observed during the preliminary experiments, they are not considered (Hamerla et al. 2013). The postulated reaction network presented in Figure 4.11 was proposed by (Kiedorf et al. 2013) and successfully applied in (Müller et al. 2013).

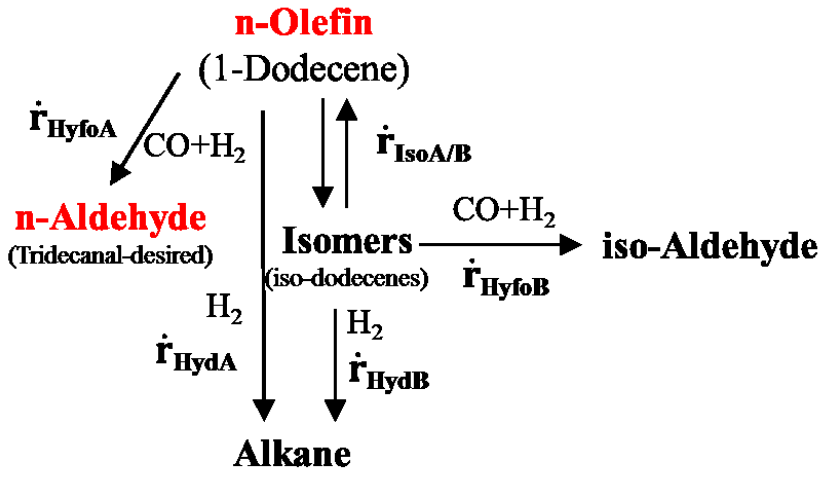


Figure 4.11 Reaction network of hydroformylation process

The reaction network includes an equilibrium reaction $\dot{r}_{IsoA/IsoB}$ between the reactant 1-dodecene and its isomers. Both 1-dodecene and iso-dodecenes react to dodecane $\dot{r}_{HydA/HydB}$ in the presence of hydrogen. The main product n-tridecanal is formed by the

[3]http://www.inprompt.tu-berlin.de/

hydroformylation $\dot{r}_{HyfoA}$of 1-dodecene with CO and H_2. Although the catalyst complex is formed with a bidentate ligand so as to raise the selectivity regarding the wanted linear product n-tridecanal and its branched isomers, small amounts of iso-aldehydes are formed represented by the hydroformylation$\dot{r}_{HyfoB}$. Preliminary experiments have not shown the creation of iso-aldehydes via 1-dodecene. Therefore, it is assumed that the iso-aldehydes are solely created from the iso-dodecenes. The component balances with respect to the amount of substance are given as follows

$$\frac{dn_{Doce}}{dt} = V_L \cdot \left(-\dot{r}_{IsoA} + \dot{r}_{IsoB} - \dot{r}_{HydA} - \dot{r}_{HyfoA}\right) \quad (4.5)$$

$$\frac{dn_{IsoDoce}}{dt} = V_L \cdot \left(\dot{r}_{IsoA} - \dot{r}_{IsoB} - \dot{r}_{HydB} - \dot{r}_{HyfoB}\right) \quad (4.6)$$

$$\frac{dn_{Doca}}{dt} = V_L \cdot \left(\dot{r}_{HydA} + \dot{r}_{HydB}\right) \quad (4.7)$$

$$\frac{dn_{TDC}}{dt} = V_L \cdot \dot{r}_{HyfoA} \quad (4.8)$$

$$\frac{dn_{IsoAld}}{dt} = V_L \cdot \dot{r}_{HyfoB} \quad (4.9)$$

The reduced mechanistic reaction kinetic model was derived based on the Wilkinson-catalyst cycle. Details of the derivation can be found in (Kiedorf et al. 2013).

$$\dot{r}_{IsoA} = \frac{k_{ref,IsoA} \cdot \exp\left(-\frac{E_{a,Iso}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) c_{Cat} \cdot c_{Doce}}{\left(1 + K_{\alpha,iso} \cdot c_{Doce}\right) \cdot \left(1 + K_{\alpha,Kat} \cdot c_{CO}\right)} \quad (4.10)$$

$$\dot{r}_{IsoB} = \frac{k_{ref,IsoB} \cdot \exp\left(-\frac{E_{a,Iso}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \cdot c_{Cat} \cdot c_{IsoDoce}}{\left(1 + K_{\alpha,iso} \cdot c_{Doce}\right) \cdot \left(1 + K_{\alpha,Kat} \cdot c_{CO}\right)} \quad (4.11)$$

$$\dot{r}_{HydA} = \frac{k_{ref,HydA} \cdot \exp\left(-\frac{E_{a,Hyd}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \cdot c_{Cat} \cdot c_{Doce} \cdot c_{H_2}}{\left(1 + K_{\alpha,hyd} \cdot c_{Doce}\right) \cdot \left(1 + K_{\alpha,Kat} \cdot c_{CO}\right)} \quad (4.12)$$

$$\dot{r}_{HydB} = \frac{k_{ref,HydB} \cdot \exp\left(-\frac{E_{a,Hyd}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) c_{Cat} \cdot c_{IsoDoce} \cdot c_{H_2}}{\left(1 + K_{\alpha,hyd} \cdot c_{Doce}\right) \cdot \left(1 + K_{\alpha,Kat} \cdot c_{CO}\right)} \quad (4.13)$$

$$\dot{r}_{HyfoA} = \frac{c_{Cat} c_{Doce} \cdot c_{H_2} \cdot c_{CO} \cdot k_{ref,HyfoA} \cdot \exp\left(-\frac{E_{a,Hyf}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right)}{\left(1 + K_{\alpha,hyfo} \cdot c_{Doce} + K_{\beta,hyfo} \cdot c_{Doce} \cdot c_{CO}\right) \cdot \left(1 + K_{\alpha,Kat} \cdot c_{CO}\right)} \quad (4.14)$$

$$\dot{r}_{HyfoB} = \frac{c_{Cat} \cdot c_{IsoDoce} \cdot c_{H_2} \cdot c_{CO} \cdot k_{ref,HyfoB} \cdot \exp\left(-\frac{E_{a,Hyf}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right)}{\left(1 + K_{\alpha,hyfo} \cdot c_{Doce} + K_{\beta,hyfo} \cdot c_{Doce} \cdot c_{CO}\right) \cdot \left(1 + K_{\alpha,Kat} \cdot c_{CO}\right)} \quad (4.15)$$

State variables of the process are the mole amount of 1-dodecene n_{Doce}, iso-dodecenes $n_{IsoDoce}$, dodecane n_{Doca}, n-tridecanal n_{TDC} and iso-aldehydes n_{IsoAld}. The temperature dependency of the reaction rates are of the centralized-temperature-form as proposed by (Buzzi-Ferraris & Manenti 2009) and (Bernas et al. 2010), which is crucially more capable for parameter estimation task than the classical Arrhenius approach.

$$c_i = \frac{n_i}{V_L}, \qquad i: Doce, IsoDoce, Doca, TDC, IsoAld \quad (4.16)$$

$$c_j = \frac{p_j}{H_{0,j} \exp\left(\frac{H_{ads,j}}{RT}\right)}, \qquad j: H_2, CO \quad (4.17)$$

The concentrations of CO and H_2 in the liquid phase are given as function of the partial pressures (4.17) via the Van't Hoff equation with respect to the adsorption enthalpy $H_{ads,j}$ (Smith & Harvey 2007) while assuming that the pressure dependency is negligible for the operating range from 5 to 15 bar and that the mass transfer from the gas phase into the liquid phase is much faster than the reaction rates. The process parameters to be examined are the frequency factors $kref_{IsoA}$, $kref_{HydA}$, $kref_{HyfoA}$, the activation energies $E_{a,Iso}$, $E_{a,Hyd}$, $E_{a,Hyfo}$ and the mechanistic parameters $K_{\alpha,Hyfo}$, $K_{\alpha,Kat}$. Furthermore, the process can be influenced by three control variables: the partial pressures p_{CO/H_2} and the reactor temperature T. The remaining degrees of freedom are grouped together as fixed model constants and listed in Table 33. The aim of this example is to show the full capability of the OED without any technical limitation regarding the implementation of the optimized control trajectories. Therefore the real reactor is simulated by the model equations (4.5) to (4.17) with the reference parameter set (Table 34) which was obtained by more than 200 single experiments without design. According to the experimental setup, the mole amount of 1-dodecene n_{Doce}, iso-dodecenes $n_{IsoDoce}$, dodecane n_{Doca} and n-tridecanal n_{TDC} can be measured through the running experiment with a measurement variance of 0.0001 mole. Furthermore the process constraints on the initial conditions and the control variables are:

$$0.01\ \text{mole} \leq n_{0,\text{Doce}} \leq 0.05\ \text{mole}$$
$$5\ \text{bar} \leq p_{CO/H_2}(t) \leq 15\ \text{bar} \qquad (4.18)$$
$$368\ \text{K} \leq T(t) \leq 388\ \text{K}$$

4.2.1 Factorial Design

First, the method of factorial design is applied to the experimental design task so as to emphasize the high experimental effort related with the conventional approach. Then the OED approach is applied to the same task so as to show how much work can be reduced at even more precisely estimated parameters. For the conventional approach, a 2k-design is applied to the partial pressures p_{CO/H_2} leading to 2^2 combinations and the 3^k-design is applied to the initial mole amount of the substrate 1-dodecene $n_{0,\text{Doce}}$ and the reactor temperatureT leading to 3^2 combinations as seen in Table 20. This results in an experiment design with 36 experiments.

Table 20 Design conditions - hydroformylation

	LOW	MEDIUM	HIGH
p_{CO/H_2}	5 bar	–	15 bar
$n_{0,\text{Doce}}$	0.01 mole	0.03 mole	0.05 mole
T	368 K	378 K	388 K

The parameter estimation of the reaction kinetics with mechanistic parameters was carried out by three subsequent steps.

1) Pre-estimation of the mechanistic parameters and the frequency factors with the experiments at the reference temperature T_{ref}
2) Pre-estimation of the activation energy with experiments at different temperature levels except the reference temperature $T_{ref}$and fixed pre-estimated mechanistic parameters and frequency factors
3) Estimation of all parameters with all experiment data

The idea behind step one is quite simple but very effective. The exponential term of all reaction rates becomes one at the reference temperature and thus the influence of the activation energies is canceled out. Since parameter estimation problems with the

conventional Arrhenius term are extremely stiff, where the condition number of the Jacobian is in general of 10^{12} magnitude or even more, this reformulation generally reduces the magnitude of the condition number down to 10^{6} or even less (Buzzi-Ferraris & Manenti 2009). The corresponding experiments of the factorial design for step one are listed in Table 35 from number 13 to 24 .For all experiments, the experiment duration is fixed to 90 minutes, the sampling number is limited to 8 samplings and the sampling policy is designed as a typical one of chemists for an estimation of kinetic parameters, where more samplings are taken at the beginning and less at the end of the experiment as given in Table 21 and illustrated in Figure 4.12.

Table 21 Hydroformylation - factorial design sampling policy

experiment duration	90 min
sampling vector [min]	$[1, 3, 8, 15, 30, 45, 60, 90]$

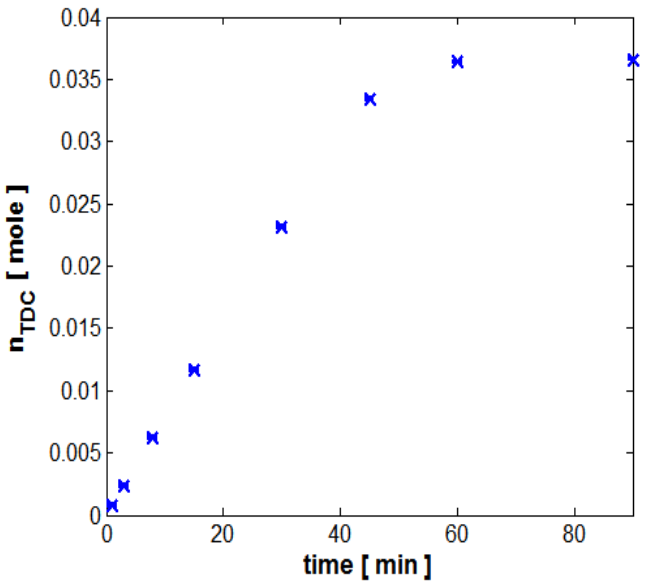


Figure 4.12 Sampling policy of chemists for kinetic parameter estimation

The parameter estimation results of the factorial design for step one are listed in Table 22. The frequency factors and the kinetic parameters could be estimated relatively near to the true parameters due to the “deactivated” effect of the activation energies.

Table 22 Parameter estimation results – Factorial design step one

	Initial guess	Final value	confidence interval - 99%	Reference value
$kref_{IsoA}$	100	7.7978	2.549	9.3912
$kref_{HydA}$	100	657.77	267.5	798.34
$kref_{HyfoA}$	100	22629	14300	27495
$K_{\alpha,Hyfo}$	1	21.944	11.91	22.376
$K_{\alpha,Kat}$	1	74.392	30.33	90.917

For step two, the pre-estimated parameters of step one are fixed. The corresponding experiments for the estimation of the high nonlinear activation energies are listed in Table 35 from number 1 to 12 and 25 to 36.The initial values for the activation energies are partially guessed values and partially based on published pre-works regarding the hydroformylation of 1-dodecene (Bhanage et al. 1997) and (Koeken et al. 2011). The estimation results of this step are listed in Table 23.

Table 23 Parameter estimation results – Factorial design step two

	Initial guess	Final value	confidence interval - 99%	Reference value
$E_{a,Iso}$	30000	91897	6917	93245
$E_{a,Hyd}$	45000	24773	2186	30820
$E_{a,Hyfo}$	60000	67048	3354	71871

The estimation results are also reasonable in this case because of the well pre-estimated parameters from step one. Finally, for step three, all pre-estimated parameters are opened again with their pre-estimated values as initial guess for the final estimation step based on the data of all 36 experiments. The estimation results of the last step are listed in Table 24.

.

Table 24 Parameter estimation results – Factorial design step three

	Initial guess	Final value	confidence interval - 99%	Reference value
$\mathrm{kref_{IsoA}}$	7.7978	13.009	2.692	9.3912
$\mathrm{kref_{HydA}}$	657.77	1110.4	268.4	798.34
$\mathrm{kref_{HyfoA}}$	22629	36777	11740	27495
$\mathrm{K_{\alpha,Hyfo}}$	21.944	20.218	4.473	22.376
$\mathrm{K_{\alpha,Kat}}$	74.392	131.75	31.82	90.917
$\mathrm{E_{a,Iso}}$	91897	87380	3297	93245
$\mathrm{E_{a,Hyd}}$	24773	32963	7462	30820
$\mathrm{E_{a,Hyfo}}$	67048	69259	2244	71871

Although the estimated parameter values are near the true parameter values and the corresponding confidence intervals are relatively small, the experiment effort was very high. Additionally to the totally 54 hours of the 36 experiments, one must also consider the subsequent treatment of the corresponding 288 GC-samplings, which is even more time consuming and costs lots of resources. Therefore, the goal of the optimal experimental design approach primarily aims at reducing the experimental effort as presented in the following section.

4.2.2 Optimal experimental design

In contrast to the factorial design, the process is addressed not only by discrete decisions but also dynamic control strategies as shown in Table 25.

Table 25 Control strategies - OED

$\mathrm{p_{CO/H_2}(t)}$	dynamic zero-order control
$\mathrm{T(t)}$	dynamic first-order control
$\mathrm{n_{Doce}(t=0)}$	discrete decision

Additionally, the following constraints are included in the optimization formulation:

Table 26 Additional limitations - OED

Time limitation	$t_{end} = 60$ min
Number of samplings	8

The parameter estimation strategy with three steps as presented in the previous section is also used for the optimal experimental design leading to three experiments as summarized in Figure 4.13. The OED problem is implemented via the global optimizer BARON and solved with respect to the A-optimal criterion.

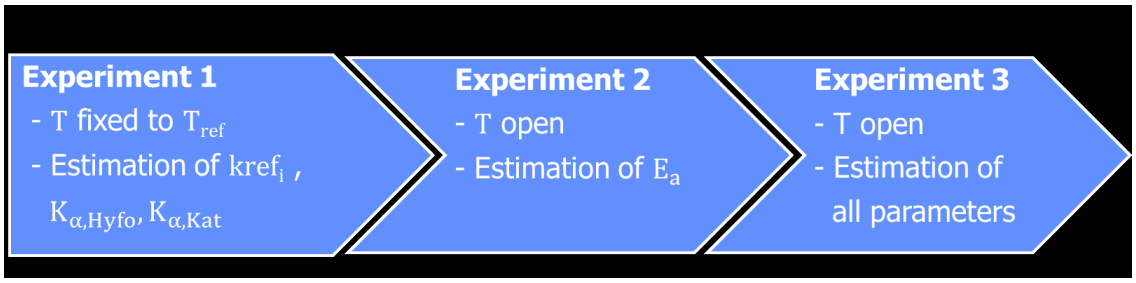


Figure 4.13 Parameter estimation strategy

In the first experiment the rector temperature is kept constant at $T = T_{ref} = 378K$. The optimized initial condition for the 1-dodecene concentration hits the upper bound with $n_{Doce,0} = 0.05$ mole. The optimized trajectories for the CO- and H_2- partial pressure are given in Figure 4.14 and Figure 4.15 respectively.

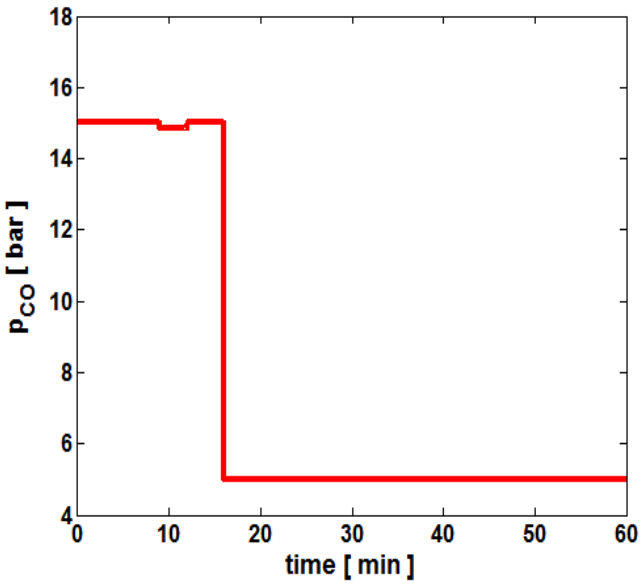


Figure 4.14 Experiment 1, optimized CO-partial pressure

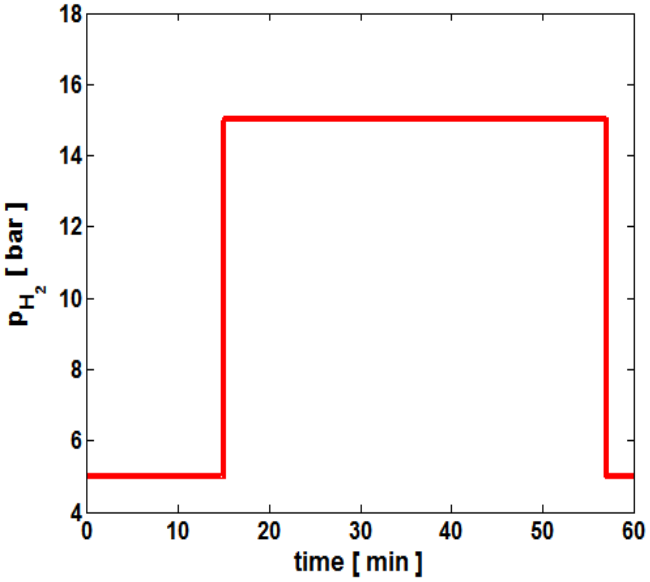


Figure 4.15 Experiment 1, optimized H_2-partial pressure

Since the H_2 partial pressure enters the model linearly (4.17), its optimal solution has to result in a bang-bang solution. This future does not apply for the temperature and the CO- partial pressure since they are nonlinear in the model equations. The OASE for 1-dodecene and TDC are shown in Figure 4.16 and Figure 4.17 and for iso-dodecenes and dodecane in Figure 6.15 and Figure 6.16 respectively.

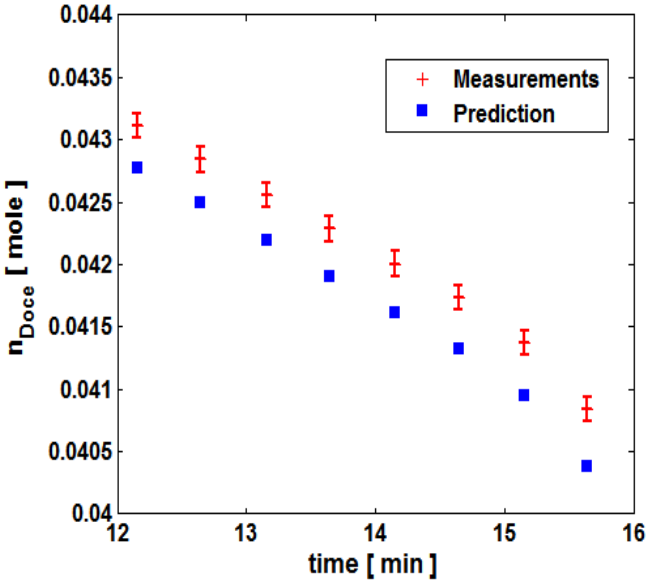


Figure 4.16 Experiment 1, OASE for 1-dodecene

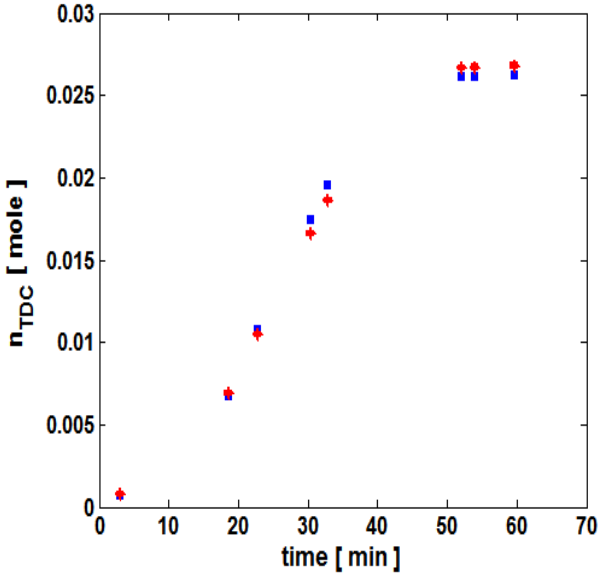


Figure 4.17 Experiment 1, OASE for TDC

The estimated parameters based on the data of experiment 1 are listed in Table 27. After the first experiment, the estimated parameter values are already relatively near to the true parameters. However their confidence intervals are still very large because of the small number of sampling data.

Table 27 Parameter estimation results – OED step one

	Initial guess	Final value	confidence interval - 99%	Reference value
$kref_{IsoA}$	100	18.242	57.47	9.3912
$kref_{HydA}$	100	1632.3	6005	798.34
$kref_{HyfoA}$	100	31134	$1.58 * 10^5$	27495
$K_{\alpha,Hyfo}$	1	9.9153	35.04	22.376
$K_{\alpha,Kat}$	1	188.23	651.1	90.917

In the second experiment the temperature is dynamically controlled so as to capture the temperature dependency of the process.

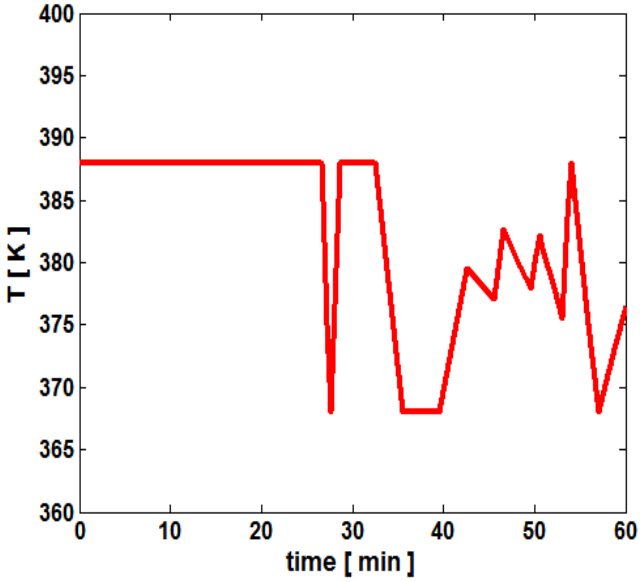


Figure 4.18 Experiment 2, reactor temperature

The optimized temperature trajectory as seen in Figure 4.18 shows the optimized dynamical perturbations of the process temperature. Since the activation energies are directly coupled, changes in the temperature directly affect the corresponding reaction rates and sensitivities. The corresponding optimized trajectories for the CO- and H_2-partial pressure are given in Figure 4.19 and Figure 4.20 respectively. From the chemical point of view, three main dynamic behaviors can be observed. The first dynamic trend starts from the beginning of the experiment and continues for 25 minutes. During this period all control variables, which are the temperature and the partial pressures of CO and H_2 are at their upper bounds. This can be explained with an optimal condition for the main hydroformylation reaction. The second dynamic trend can be observed from the 25^{th} to the 54^{th} minute. The partial pressure of H_2 is reduced to its lower bound whereas the perturbations of the temperature and the partial pressure of CO take action decisively. With the reduced hydrogen concentration in the gas phase the isomerization reaction is preferred, hence its activation energy and the corresponding sensitivities can be addressed more precisely.

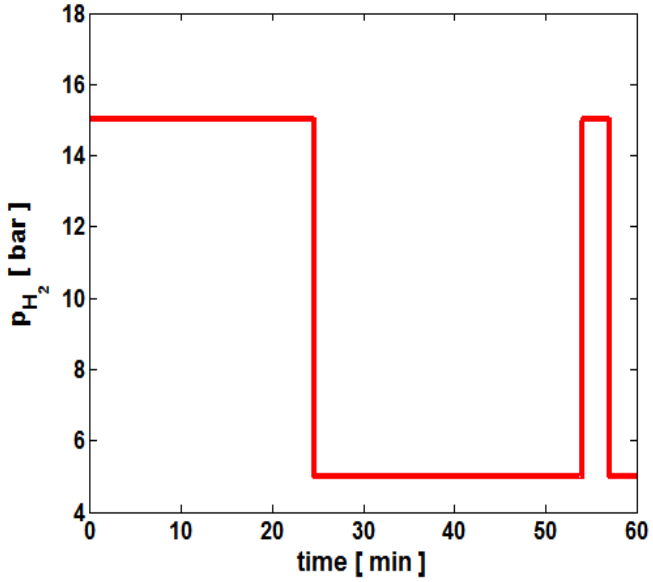


Figure 4.19 Experiment 2, optimized CO-partial pressure

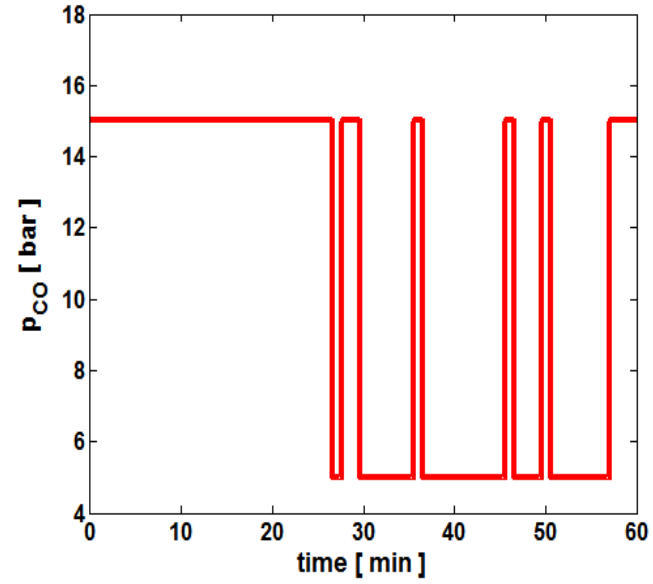


Figure 4.20 Experiment 2, optimized H_2-partial pressure

The last dynamic trend starts from the 54th minute and reaches the end of the experiment. With the increase of H_2 to its upper bound and the parallel reduction of CO to its lower bound the hydrogenation reaction is now preferred, hence the corresponding activation energy and sensitivities can be fully addressed. The OASE for the 1-dodecene and the TDC mole amounts are shown in Figure 4.21, Figure 4.22 and Figure 6.17, Figure 6.18 for iso-dodecenes and dodecane respectively.

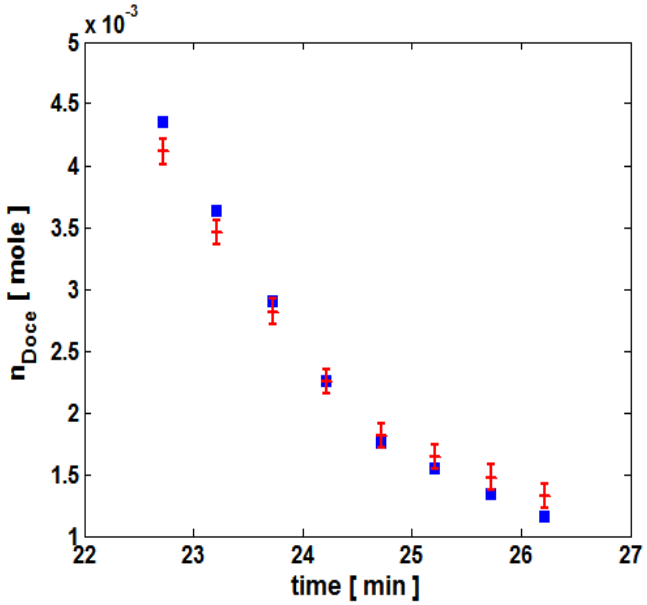


Figure 4.21 Experiment 2, OASE for 1-dodecene

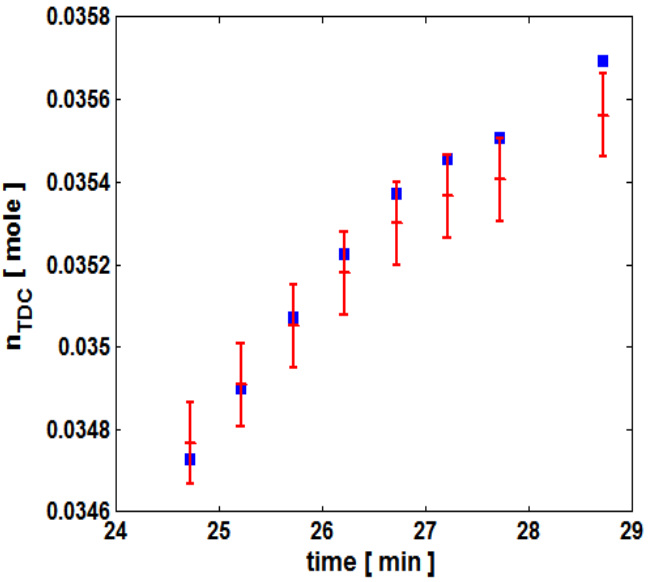


Figure 4.22 Experiment 2, OASE for TDC

The parameter estimation results for the activation energies based on the sampling data of experiment 2 are listed in Table 28.

Table 28 Parameter estimation results – OED step two

	Initial guess	Final value	confidence interval - 99%	Reference value
$E_{a,Iso}$	30000	88451	14410	93245
$E_{a,Hyd}$	45000	39079	30630	30820
$E_{a,Hyfo}$	60000	71578	3700	71871

Finally all pre-estimated parameter are opened again for the estimation in experiment 3. Since all parameter are already moved near the reference parameter values, the optimized temperature control trajectory is not as aggressive as in experiment 2 as seen in Figure 4.23.

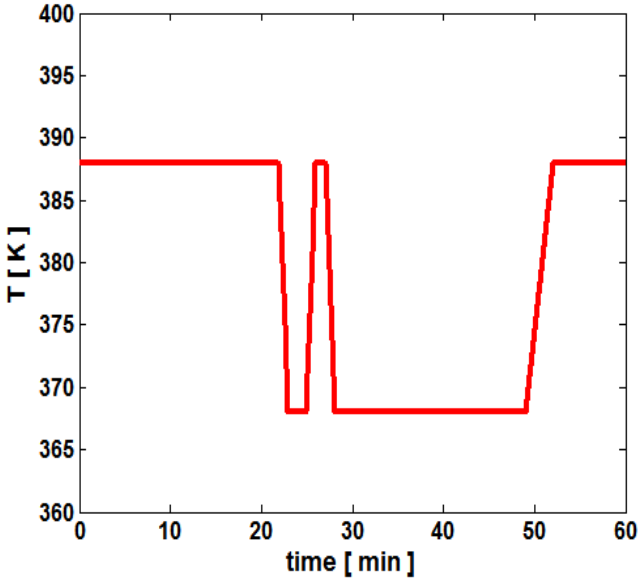


Figure 4.23Experiment 3, reactor temperature

This also applies for the optimized trajectories of the CO- and H_2- partial pressures as shown in Figure 4.25 and Figure 4.24. Analogously to experiment 2, it can be also observed that the hydroformylation reaction is preferred in the first part of the experiment. One short perturbation around the 22nd and the 25th minute addresses one more time the isomerization reaction. With the total reduction of CO, the third and last part of the experiment clearly addresses the hydrogenation reaction.

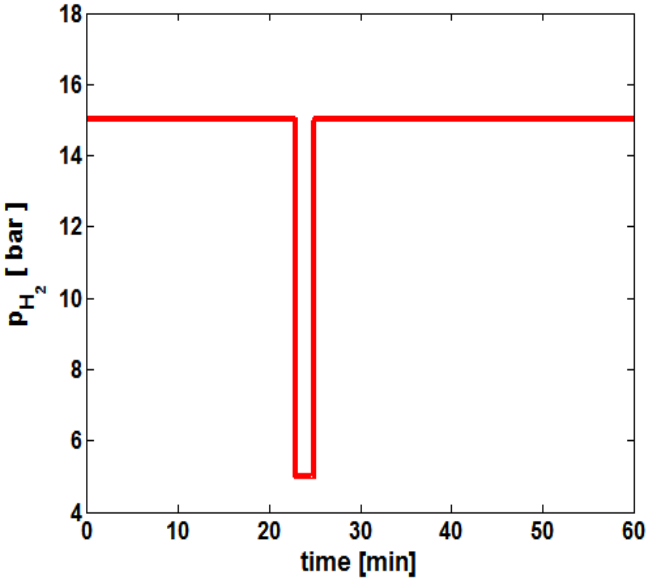


Figure 4.24 Experiment 3, optimized H_2-partial pressure

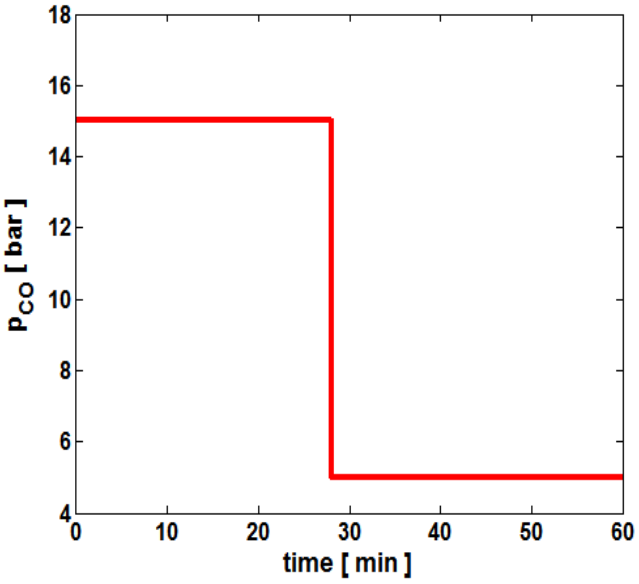


Figure 4.25 Experiment 3, optimized CO-partial pressure

The OASE of experiment 3 for the 1-dodecene and the TDC mole amounts are shown in Figure 4.26, Figure 4.27 and Figure 6.19, Figure 6.20 for iso-dodecenes and dodecane respectively.

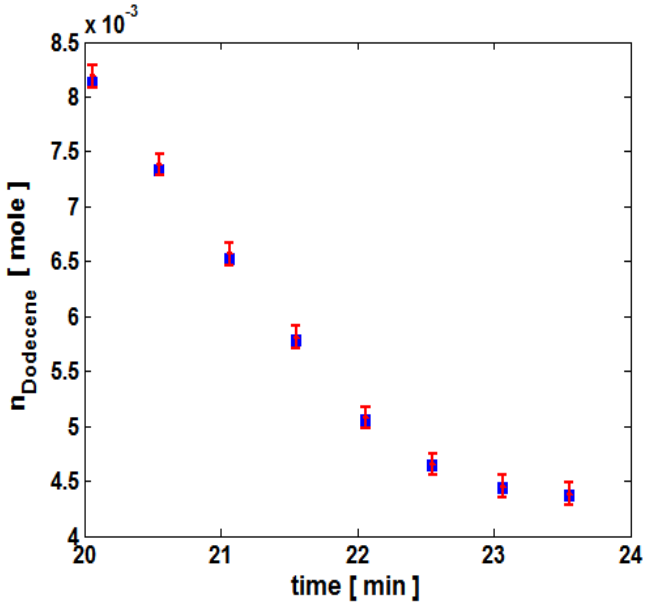


Figure 4.26 Experiment 3, OASE for 1-dodecene

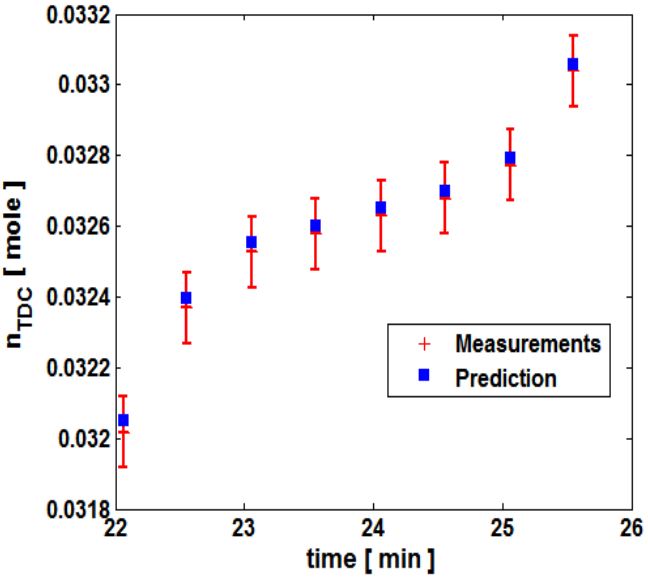


Figure 4.27 Experiment 3, OASE for TDC

Taking the sampling information of all three experiments together, the final parameter estimation step leads to the results which are listed in Table 29.

Table 29 Parameter estimation results – OED step three

	Initial guess	Final value	confidence interval - 99%	Reference value
$kref_{IsoA}$	18.242	9.0776	0.5708	9.3912
$kref_{HydA}$	1632.3	769.22	58.45	798.34
$kref_{HyfoA}$	31134	24979	3721	27495
$K_{\alpha,Hyfo}$	9.9153	20.538	3.384	22.376
$K_{\alpha,Kat}$	188.23	87.753	6.271	90.917
$E_{a,Iso}$	88451	94098	2043	93245
$E_{a,Hyd}$	39079	27238	4081	30820
$E_{a,Hyfo}$	71578	72955	2485	71871

The estimated parameters after three optimized dynamic experiments are satisfied. Furthermore, their confidence intervals are relatively small meaning that the estimations are accurate and trustable. A comparison between the estimation results between factorial design and OED is shown in Table 30.

Table 30 Comparison estimation results – Factorial design vs. OED

	Final value Factorial	99% - conf. interval	Final value OED	99% - conf. interval	Reference value
$kref_{IsoA}$	13.009	2.692	**9.0776**	**0.5708**	**9.3912**
$kref_{HydA}$	1110.4	268.4	**769.22**	**58.45**	**798.34**
$kref_{HyfoA}$	36777	11740	**24979**	**3721**	**27495**
$K_{\alpha,Hyfo}$	20.218	4.473	**20.538**	**3.384**	**22.376**
$K_{\alpha,Kat}$	131.75	31.82	**87.753**	**6.271**	**90.917**
$E_{a,Iso}$	87380	3297	**94098**	**2043**	**93245**
$E_{a,Hyd}$	**32963**	7462	27238	**4081**	**30820**
$E_{a,Hyfo}$	69259	**2244**	**72955**	2485	**71871**

More interesting is the question of how much experimental effort could be saved if preferring the OED approach. A comparison between both methods is illustrated in Figure 4.28.

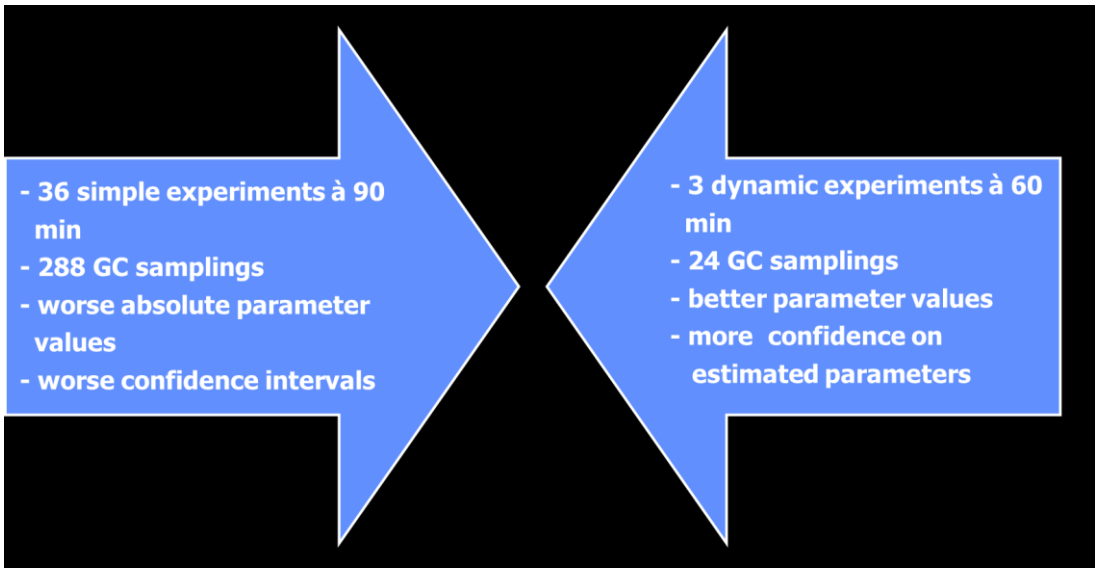


Figure 4.28 Comparison - factorial design vs. OED

In the case of the factorial design, 36 “simple” experiments were carried out, each with duration of 90 minutes, and 288 GC analyses were needed for the post treatment of the samplings. On the other hand, only three experiments à 60 minutes are needed with totally 24 GC samplings when applying the OED approach. The price in return for this is difficult dynamic experiments which have to be realized in practice. The gain is that the total time could be saved with a factor of 20 and the total samplings number could be saved with a factor of 10. More important are the estimation results. Here the OED approach clearly is superior to the factorial design approach not only with respect to the nominal estimated parameter values but also to the corresponding confidence intervals.

5 Summary and outlook

The proposed simultaneous optimization approach to OED combines two powerful methodologies together and represents the state-of-the-art method when it comes down to solve design of experiments tasks regarding precise parameter estimation. First, it inherits the formulation of an optimal control problem from the OED methodology, thus, model structure, parameters and the time dependent character of dynamic processes can be fully exploited. Secondly, the transcription of the optimal control formulation to a NLP problem covers all demands of an advanced OED formulation and it is appropriate to be solved efficiently by global optimization algorithms. The obtained results outmatch the results from sequential optimization approaches which in general only reach local optimality. Furthermore, old burdens from sequential optimization approaches like problems with instable systems, overloaded computational effort in case of many parameters and controls, inefficient Jacobian handling, high-index path constraints as well as singular control problems can be totally overcome, or at least be treated in a proper way with respect to the latter[4]. The application example of an instable CSTR process has shown that the proposed approach properly handled the instabilities and obtained optimal results, whereas state-of-the-art commercial tool for OED totally fails even in finding a feasible solution. The results of the second application example regarding the estimation of kinetic parameters of a complex hydroformylation process has shown that the proposed approach is clearly the superior method in comparison with the conventional factorial design strategy. Not only the experimental effort could remarkably be reduced but also better estimated parameter values and even more important better confidence intervals could be obtained.

It has been derived that the structure of the constraint derivatives of the formulated NLP problem are extremely sparse and most part of the first-order constraint derivatives can be found again in higher order derivatives. Therefore, all those terms can be reused during the calculations. It has also been shown that the initial conditions of the state variables as well as the decision interval lengths linearly enter the Lagrangian. Thus, if no additional constraints are violated, for example a limited experiment duration or more strict constraints on the state variables themselves, then optimality can only be obtained if their outputs result in a bang-bang solution. The most crucial decision

[4]Detailed analyzes and proofs are reserved for the interested reader in (Biegler 2010)

variable regarding OED is the sampling decisions which represent the core part of the OASE. The objective functions with respect to the FIM are quadratic in the sampling decisions. However, strict monotonicity is always given over the related special optimization space of $[0,1]$, thus, the optimum is always reached at reasonable solutions, namely 0 or 1. On the other hand the strict monotonicity cannot be guaranteed if the objective function is formulated with respect to the covariance matrix C_θ, hence, the presented approach is only used in connection with the FIM.

The proposed advanced OED approach, which includes variable control orders, optimization of the initial conditions and the decision interval length, strict compliance to process constraints, limitation on the experiment duration and the OASE has been implemented in MOSAIC[5], a freeware web based modeling environment for process systems engineering (Zerry et al. 2004; Kuntsche et al. 2011). The designed interface in MOSAIC allows the users to enter all specifications of the advanced OED task resulting in runnable program codes which then can be directly solved in GAMS via the global optimizer BARON (see Figure 6.21 and Figure 6.22).

There are some issues which are worth to discuss in the following for future investigations. A proper handling of the discretization error should be included to the optimization formulation. It is obvious that the upper bound of the decision interval length h_l cannot be as large as possible. Accordingly, it also cannot be answer just by heuristics, whether the lower bounds should be at least as small as the minimum possible sampling interval or smaller. This issue has to be rather investigated in connection with the approximation of the global discretization error$e(t) = z(t) - z^K(t)$ which holds for the OCFEM with Legendre roots

$$\max_{t\in[0,t_f]} \|e(t)\| \leq C \max_{l\in[1,\ldots,N_e]} \left(h_l^K \|T_l(t)\|\right) + O(h_l^{K+1})$$

Here, C denotes a constant which is of mathematical interest, and $T_l(t)$ depends only on the solution $z(t)$ and is independent of the choice of collocation interval length h_l. A formulation which allows variable decision interval lengths in the context of tracking and adapting to steep profiles has been proposed by (Vasantharajan & Biegler 1990). A theory to extend OP2 could be derived based from this idea.

[5]http://www.mosaic-modeling.de/

$$\sum h_l = t_f, \qquad h_l \geq 0, \qquad l = 1, \ldots, N_e$$

$$\tilde{C}\|T_l(\tau)\| \leq \varepsilon$$

Another problem is related to local optimality of the obtained solutions. The non-optimality of the solutions obtained by the implementation in Tomlab via SNOPT reveals a strong tendency to local minima so that gradient-based optimization algorithms generally cannot guarantee a global solution. On the other hand, the simultaneous approach is more favorable compare to sequential approaches regarding algorithms for global optimality because of its fully discretized formulation and the fact that the model nonlinearities are not increasing through the discretization scheme. The implementation in GAMS via BARON has clearly shown its superiority but is only capable for the A-optimal criterion since there is no direct way to represent the determinant or the eigenvalue of a matrix in GAMS. Therefore one important issue for future investigations is to apply the proposed approach with respect to D- and E-optimal criteria in connection with more flexible global optimization algorithms.

6 Appendix

6.1 Appendix for the biomass reactor process

Table 31 List of admissible experiments

	$c_{B,0}$	$c_{S,0}$	u_1	u_2	Admissible
1.	L	L	L	L	No
2.	L	L	L	H	Yes
3.	L	L	H	L	No
4.	L	L	H	H	No
5.	L	M	L	L	No
6.	L	M	L	H	Yes
7.	L	M	H	L	No
8.	L	M	H	H	No
9.	L	H	L	L	Yes
10.	L	H	L	H	Yes
11.	L	H	H	L	No
12.	L	H	H	H	No
13.	M	L	L	L	No
14.	M	L	L	H	Yes
15.	M	L	H	L	No
16.	M	L	H	H	No
17.	M	M	L	L	No
18.	M	M	L	H	Yes
19.	M	M	H	L	No
20.	M	M	H	H	No
21.	M	H	L	L	No

22.	M	H	L	H	Yes
23.	M	H	H	L	No
24.	M	H	H	H	No
25.	H	L	L	L	No
26.	H	L	L	H	Yes
27.	H	L	H	L	No
28.	H	L	H	H	No
29.	H	M	L	L	No
30.	H	M	L	H	Yes
31.	H	M	H	L	No
32.	H	M	H	H	No
33.	H	H	L	L	No
34.	H	H	L	H	No
35.	H	H	H	L	No
36.	H	H	H	H	No

6.2 Appendix for the Optimal Adaptive Sampling Strategy – OASE

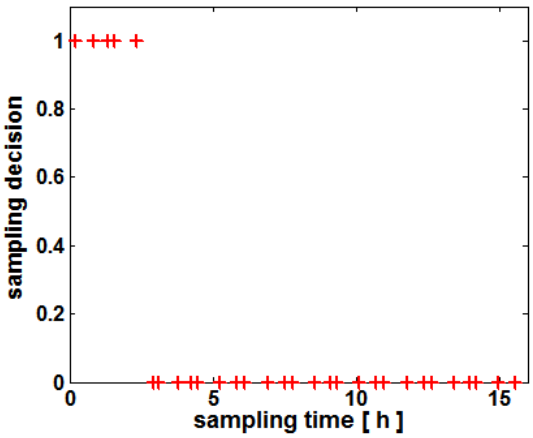


Figure 6.1 Sampling strategy - biomass concentration, OASE - SNOPT

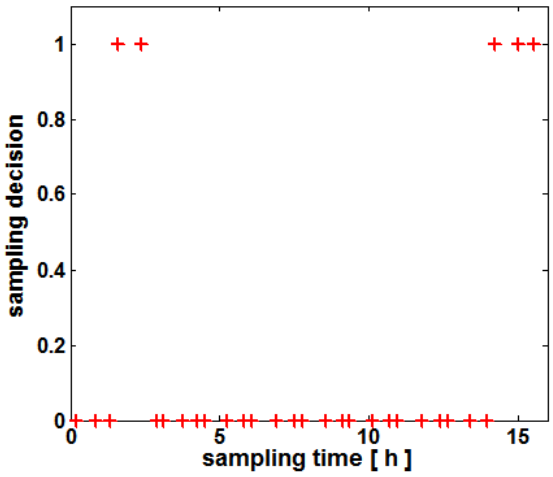


Figure 6.2 Sampling strategy - substrate concentration, OASE–SNOPT

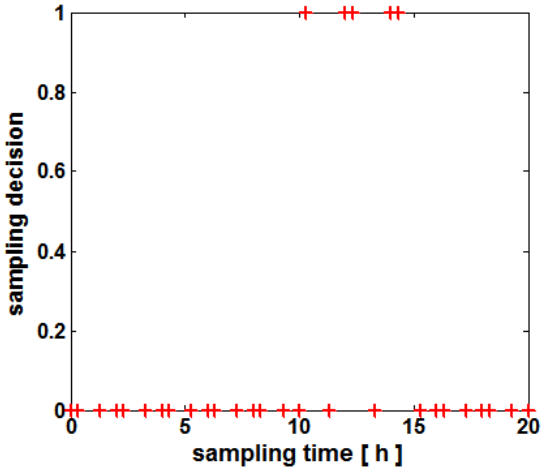


Figure 6.3 Sampling strategy - biomass concentration, OASE – BARON

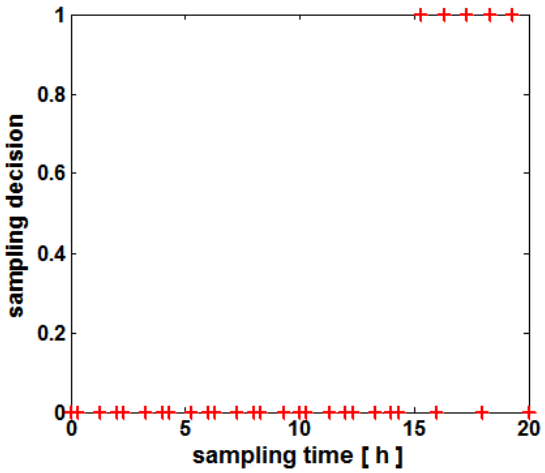


Figure 6.4 Sampling strategy – substrate concentration, OASE – BARON

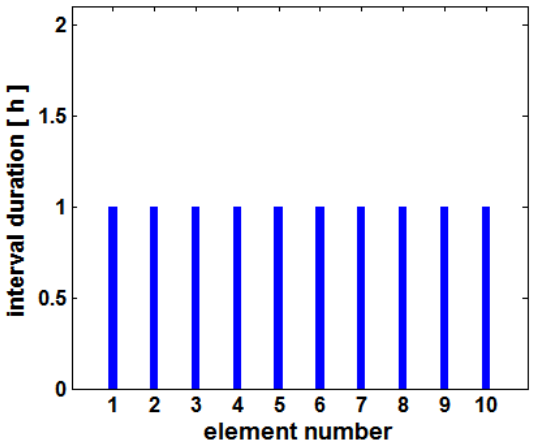


Figure 6.5 Decision intervals, D-optimal – SNOPT

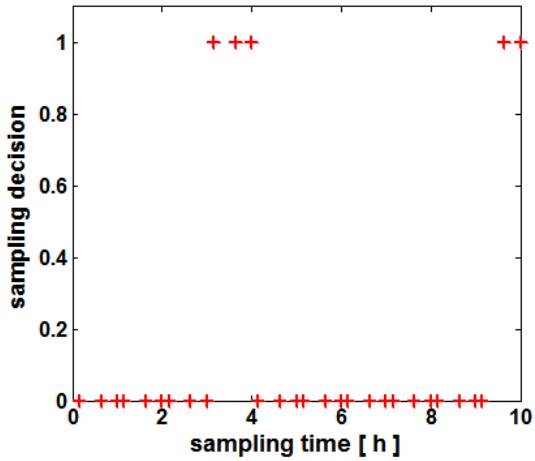


Figure 6.6 Sampling strategy - biomass concentration, D-optimal - SNOPT

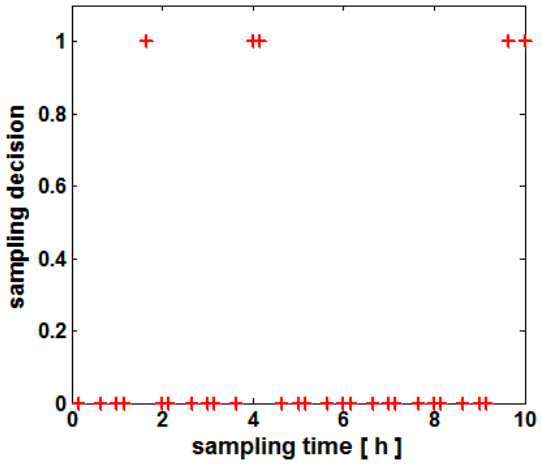


Figure 6.7 Sampling strategy – substrate concentration, D-optimal – SNOPT

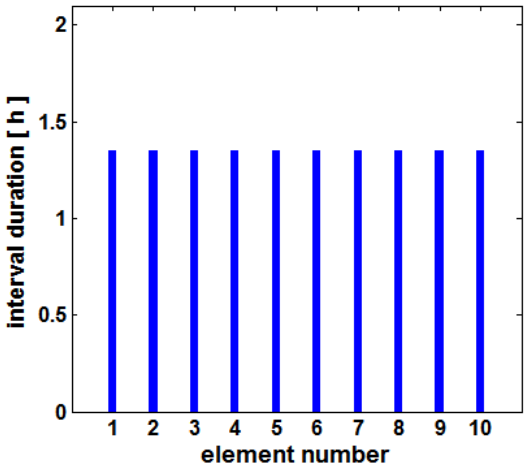


Figure 6.8 Decision intervals, E-optimal - SNOPT

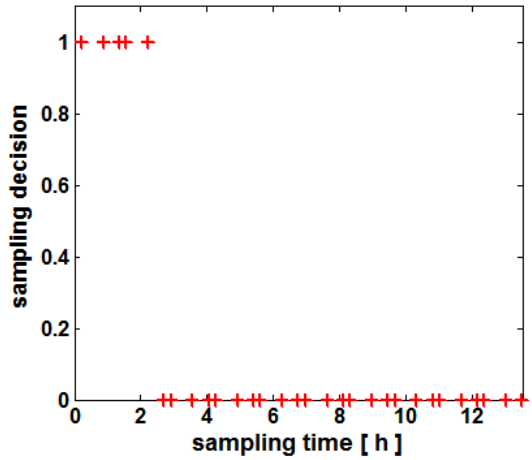


Figure 6.9 Sampling strategy - biomass concentration, E-optimal - SNOPT

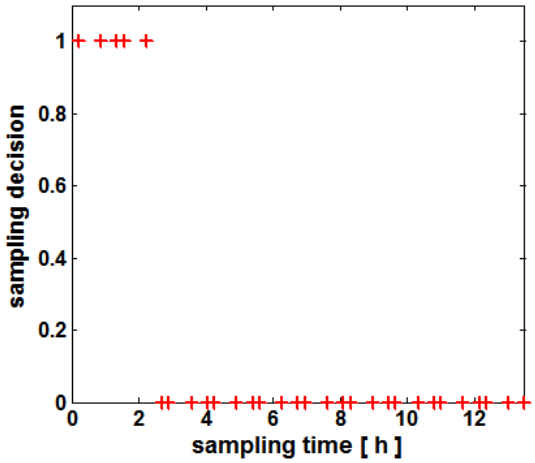


Figure 6.10 Sampling strategy – substrate concentration, E-optimal - SNOPT

6.3 Appendix for the CSTR process

Table 32 Process parameters of the CSTR-model

Parameter	Value [Unit]	Parameter	Value [Unit]
L	6.6 [dm]	ρ	$1\left[\frac{kg}{l}\right]$
F_{in}	$100\left[\frac{l}{min}\right]$	C_p	$239\left[\frac{J}{kgK}\right]$
T_{in}	350 [K]	ΔH_r	$-5\cdot 10^4\left[\frac{J}{mol}\right]$
F_{out}	$100\left[\frac{l}{min}\right]$	U	$549.36\left[\frac{J}{min\cdot dm^2K}\right]$
r	2.19 dm	$k_{r,0}$	$7.2\cdot 10^{10}\left[\frac{1}{min}\right]$
E	$72740\left[\frac{J}{mol}\right]$		

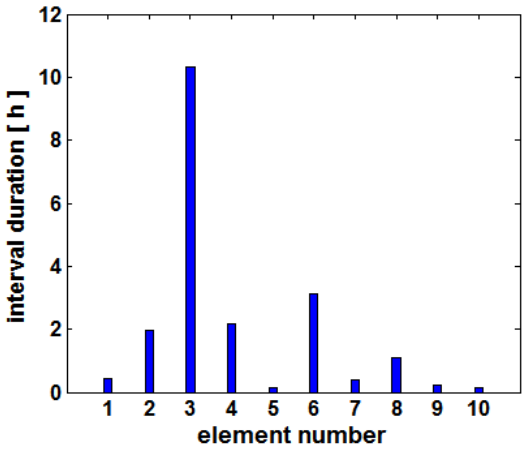


Figure 6.11 Decision intervals, sequential approach

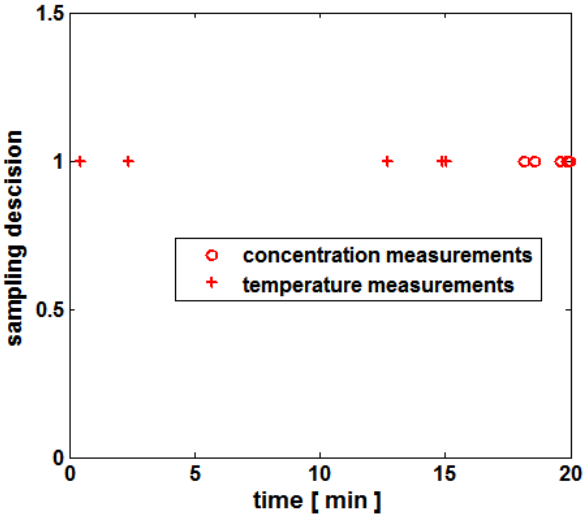


Figure 6.12 OASE, sequential approach

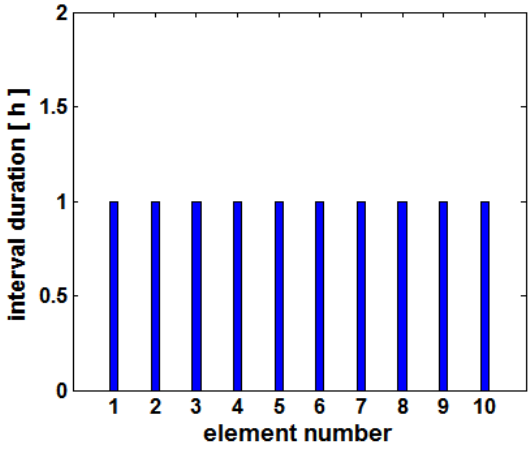


Figure 6.13 Decision intervals, simultaneous approach

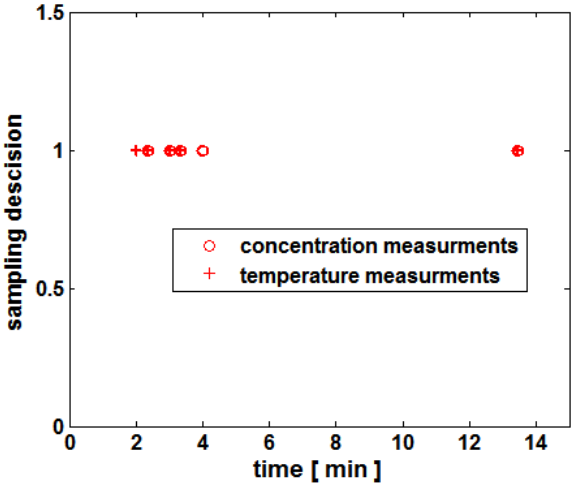
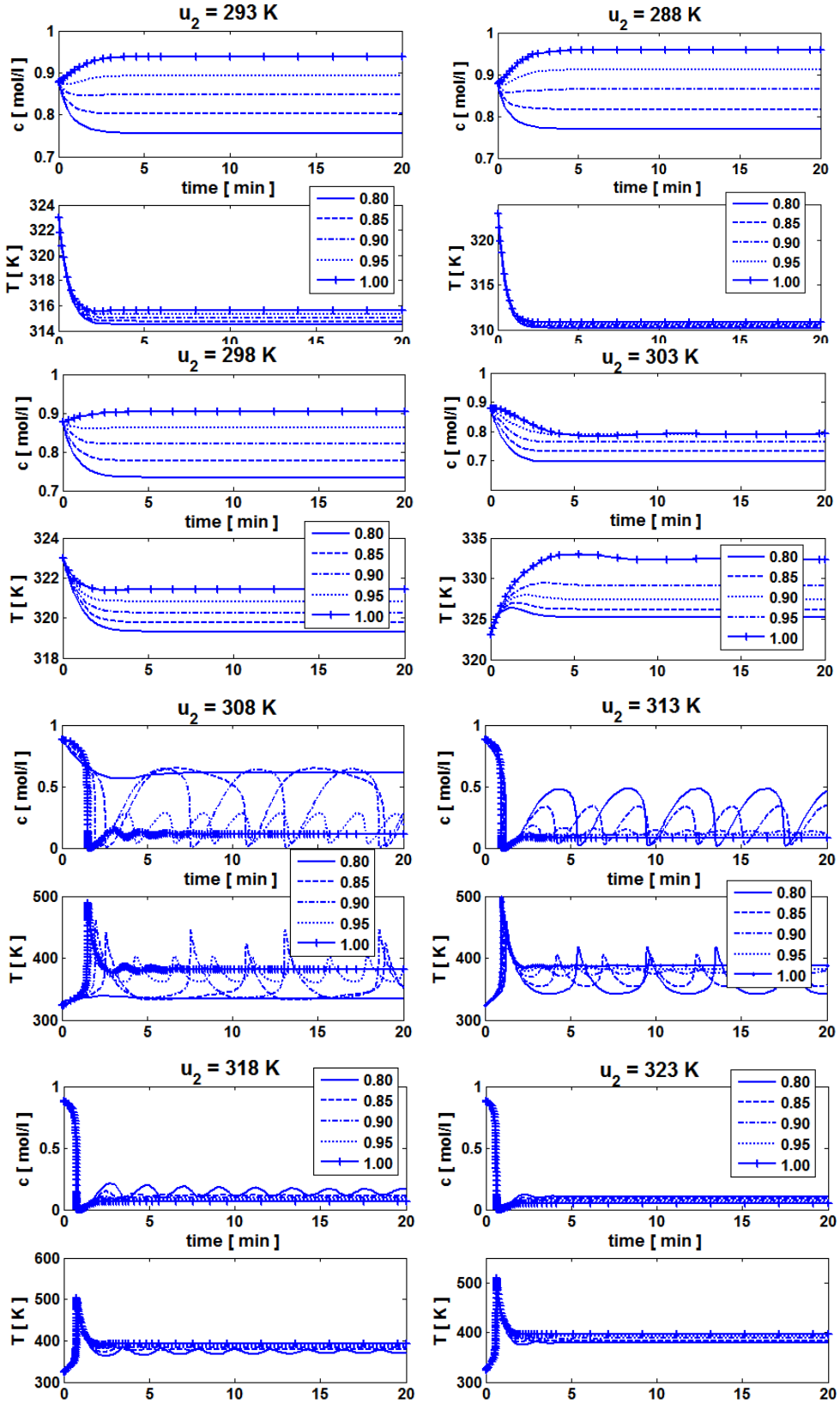
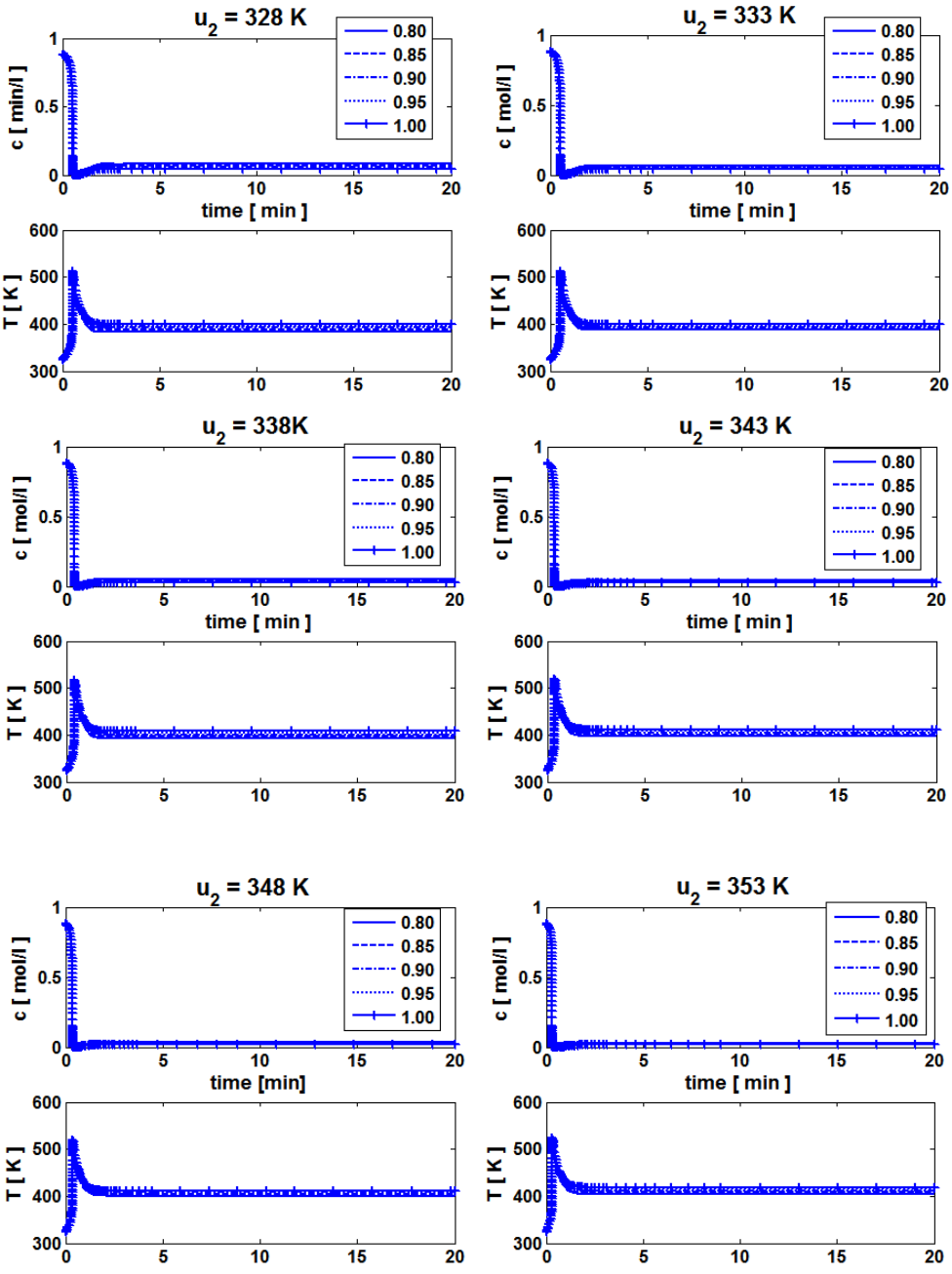


Figure 6.14 OASE, simultaneous approach


u2 = 293 K
u2 = 288 K
u2 = 298 K
u2 = 303 K
u2 = 308 K
u2 = 313 K
u2 = 318 K
u2 = 323 K
c [mol/l]
T [K]
time [min]
0.80
0.85
0.90
0.95
1.00


u2 = 328 K
u2 = 333 K
u2 = 338K
u2 = 343 K
u2 = 348 K
u2 = 353 K
0.80
0.85
0.90
0.95
1.00
c [min/l]
c [mol/l]
T [K]
time [min]
time [min]

6.4 Appendix for the hydroformylation process

Table 33 Process parameters of the hydroformylation model

Constant	Value [Unit]	Constant	Value [Unit]
$\mathrm{kref}_{\mathrm{IsoB}}$	$0.55099\left[\frac{\mathrm{l}}{\mathrm{g}}\right]$	$\mathrm{c}_{\mathrm{cat}}$	$0.0242\left[\frac{\mathrm{g}}{\mathrm{l}}\right]$
$\mathrm{kref}_{\mathrm{HydB}}$	$16.851\left[\frac{\mathrm{l}^2}{\mathrm{g}\cdot\mathrm{mol}}\right]$	V_{L}	$0.045[\,\mathrm{l}\,]$
$\mathrm{kref}_{\mathrm{HyfoB}}$	$85.822\left[\frac{\mathrm{l}^3}{\mathrm{g}\cdot\mathrm{mol}^2}\right]$	$\mathrm{H}_{0,\mathrm{H2}}$	$13.4\left[\frac{\mathrm{bar}\cdot\mathrm{l}}{\mathrm{mol}}\right]$
$\mathrm{K}_{\alpha,\mathrm{Iso}}$	$1.3597\left[\frac{\mathrm{l}}{\mathrm{mol}}\right]$	$\mathrm{H}_{0,\mathrm{CO}}$	$254.65\left[\frac{\mathrm{bar}\cdot\mathrm{l}}{\mathrm{mol}}\right]$
$\mathrm{K}_{\alpha,\mathrm{Hyd}}$	$55.719\left[\frac{\mathrm{l}}{\mathrm{mol}}\right]$	$\mathrm{H}_{\mathrm{ads,H2}}$	$8122.46\ \left[\frac{\mathrm{J}}{\mathrm{mol}}\right]$
$\mathrm{K}_{\beta,\mathrm{Hyfo}}$	$65.833\left[\frac{\mathrm{l}}{\mathrm{mol}}\right]$	$\mathrm{H}_{\mathrm{ads,CO}}$	$-2877.42\ \left[\frac{\mathrm{J}}{\mathrm{mol}}\right]$
R	$8.314\left[\frac{\mathrm{J}}{(\mathrm{mol\ K})}\right]$	$\mathrm{T}_{\mathrm{ref}}$	378 [K]

Table 34 Reference parameter set for the hydroformylation process

Parameter	Value [Unit]	Parameter	Value [Unit]
$\mathrm{kref}_{\mathrm{IsoA}}$	$9.3912\left[\frac{\mathrm{l}}{\mathrm{g}}\right]$	$\mathrm{K}_{\alpha,\mathrm{Hyfo}}$	$22.376\left[\frac{\mathrm{l}}{\mathrm{mol}}\right]$
$\mathrm{kref}_{\mathrm{HydA}}$	$798.34\left[\frac{\mathrm{l}^2}{\mathrm{g}\cdot\mathrm{mol}}\right]$	$\mathrm{K}_{\alpha,\mathrm{Kat}}$	$90.917\left[\frac{\mathrm{l}}{\mathrm{mol}}\right]$
$\mathrm{kref}_{\mathrm{HyfoA}}$	$27495\left[\frac{\mathrm{l}^3}{\mathrm{g}\cdot\mathrm{mol}^2}\right]$	$\mathrm{Ea}_{\mathrm{Iso}}$	$93245\left[\frac{\mathrm{J}}{\mathrm{mol}}\right]$
$\mathrm{K}_{\alpha,\mathrm{Iso}}$	$1.3597\left[\frac{\mathrm{l}}{\mathrm{mol}}\right]$	$\mathrm{Ea}_{\mathrm{Hyd}}$	$30820\left[\frac{\mathrm{J}}{\mathrm{mol}}\right]$
$\mathrm{K}_{\alpha,\mathrm{Hyd}}$	$55.719\left[\frac{\mathrm{l}}{\mathrm{mol}}\right]$	$\mathrm{Ea}_{\mathrm{Hyfo}}$	$71871\left[\frac{\mathrm{J}}{\mathrm{mol}}\right]$

Table 35 Factorial design – hydroformylation process

	T	$n_{0,Doce}$	p_{CO}	p_{H_2}
1.	L	L	L	L
2.	L	L	L	H
3.	L	L	H	L
4.	L	L	H	H
5.	L	M	L	L
6.	L	M	L	H
7.	L	M	H	L
8.	L	M	H	H
9.	L	H	L	L
10.	L	H	L	H
11.	L	H	H	L
12.	L	H	H	H
13.	M	L	L	L
14.	M	L	L	H
15.	M	L	H	L
16.	M	L	H	H
17.	M	M	L	L
18.	M	M	L	H
19.	M	M	H	L
20.	M	M	H	H
21.	M	H	L	L
22.	M	H	L	H
23.	M	H	H	L
24.	M	H	H	H

25.	H	L	L	L
26.	H	L	L	H
27.	H	L	H	L
28.	H	L	H	H
29.	H	M	L	L
30.	H	M	L	H
31.	H	M	H	L
32.	H	M	H	H
33.	H	H	L	L
34.	H	H	L	H
35.	H	H	H	L
36.	H	H	H	H

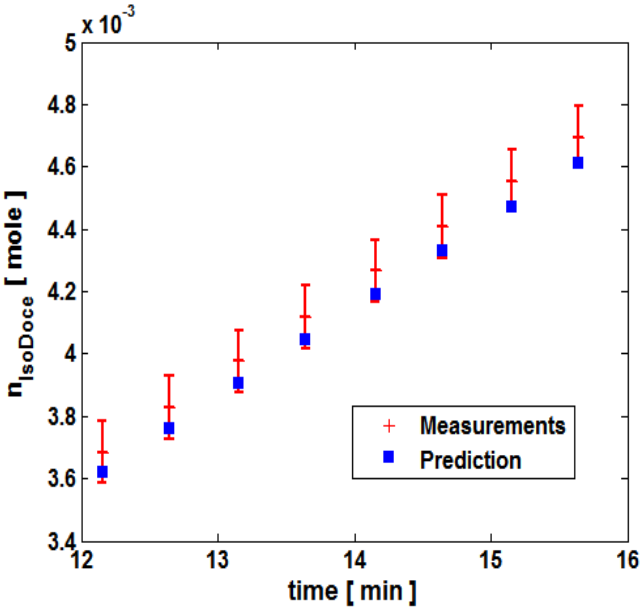


Figure 6.15 Experiment 1, OASE for iso-dodecenes

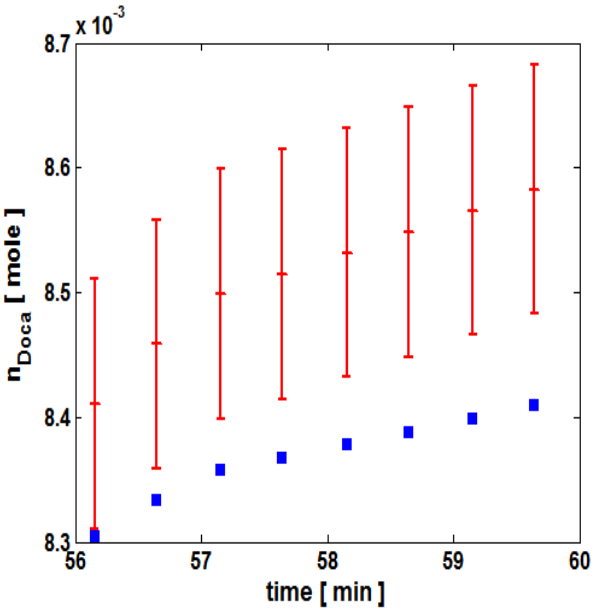


Figure 6.16 Experiment 1, OASE for dodecane

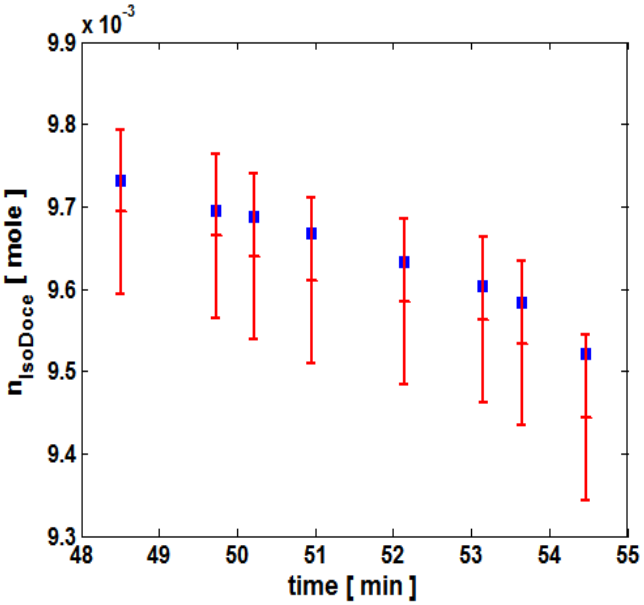


Figure 6.17 Experiment 2, OASE for iso-dodecenes

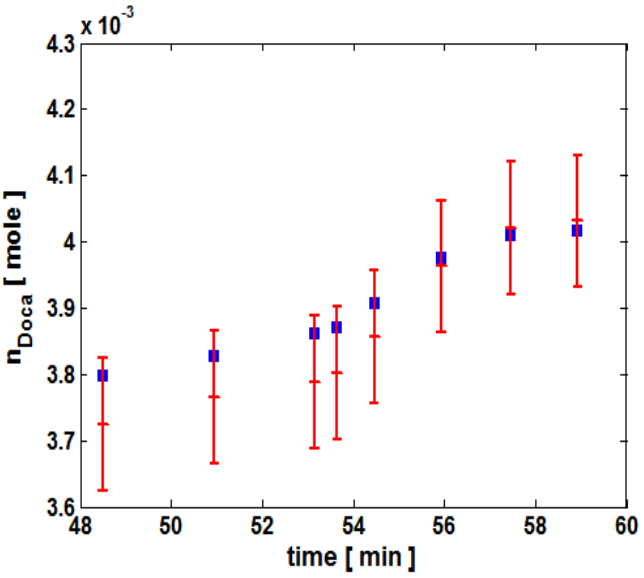


Figure 6.18 Experiment 2, OASE for dodecane

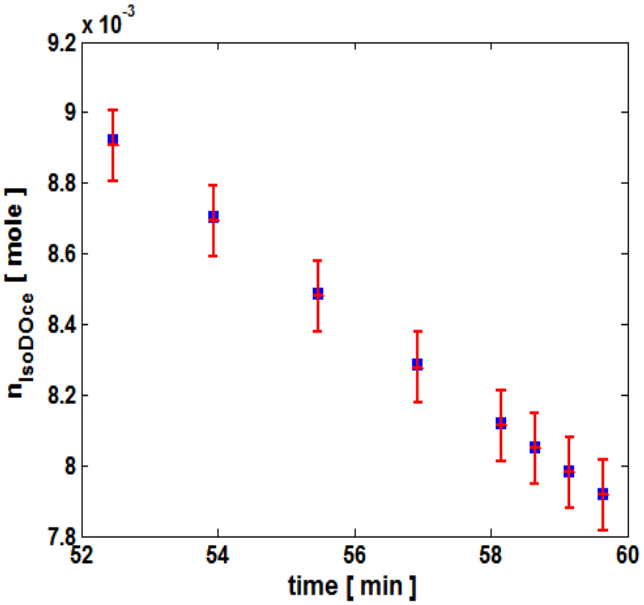


Figure 6.19 Experiment 3, OASE for iso-dodecenes

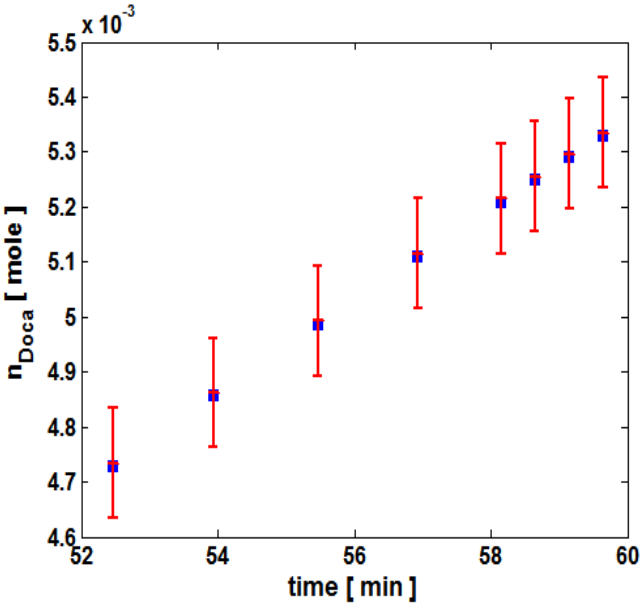


Figure 6.20 Experiment 3, OASE for dodecane

6.5 Appendix – advanced OED in MOSAIC

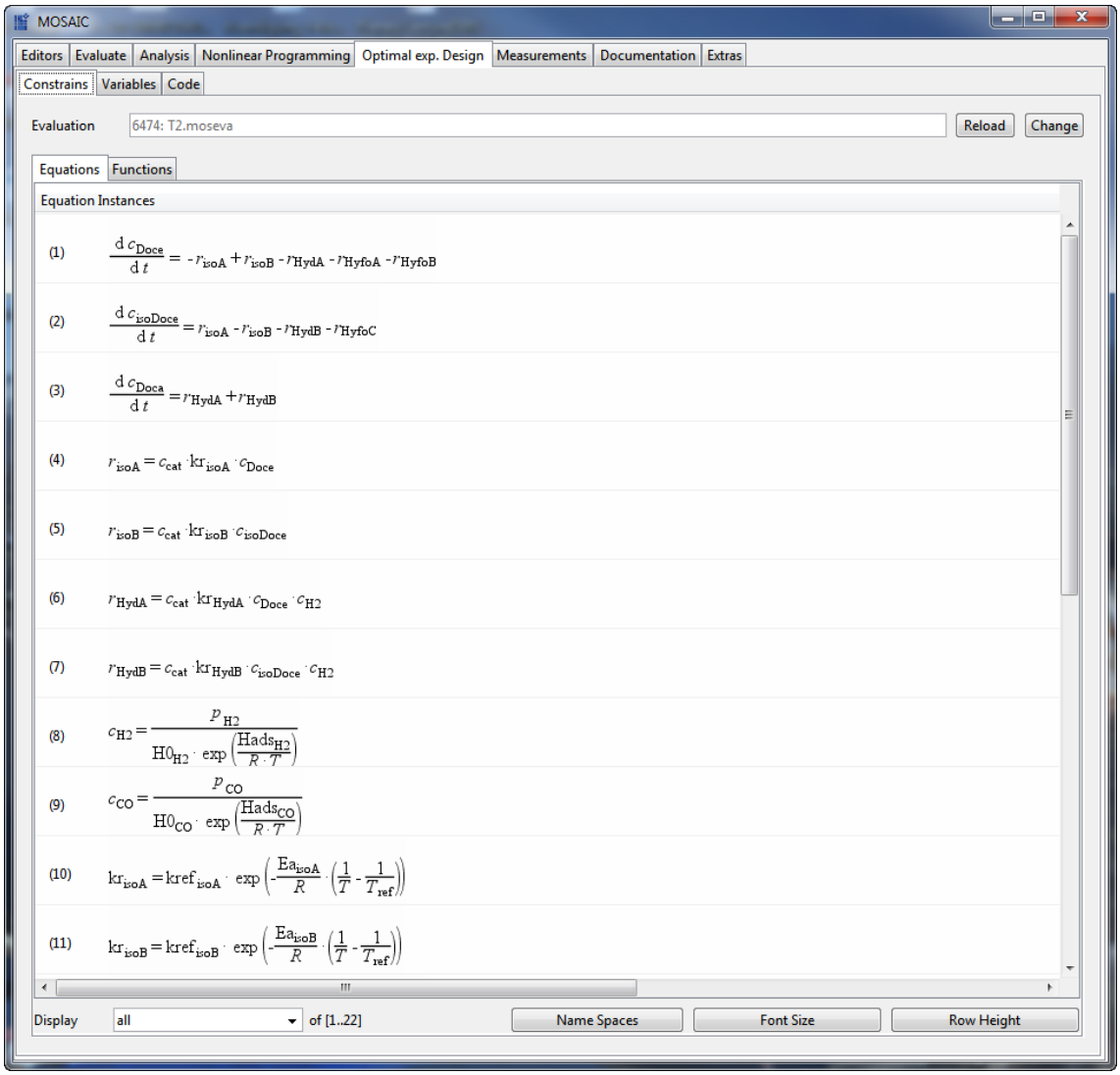


Figure 6.21 MOSAIC – modeling environment for process systems engineering

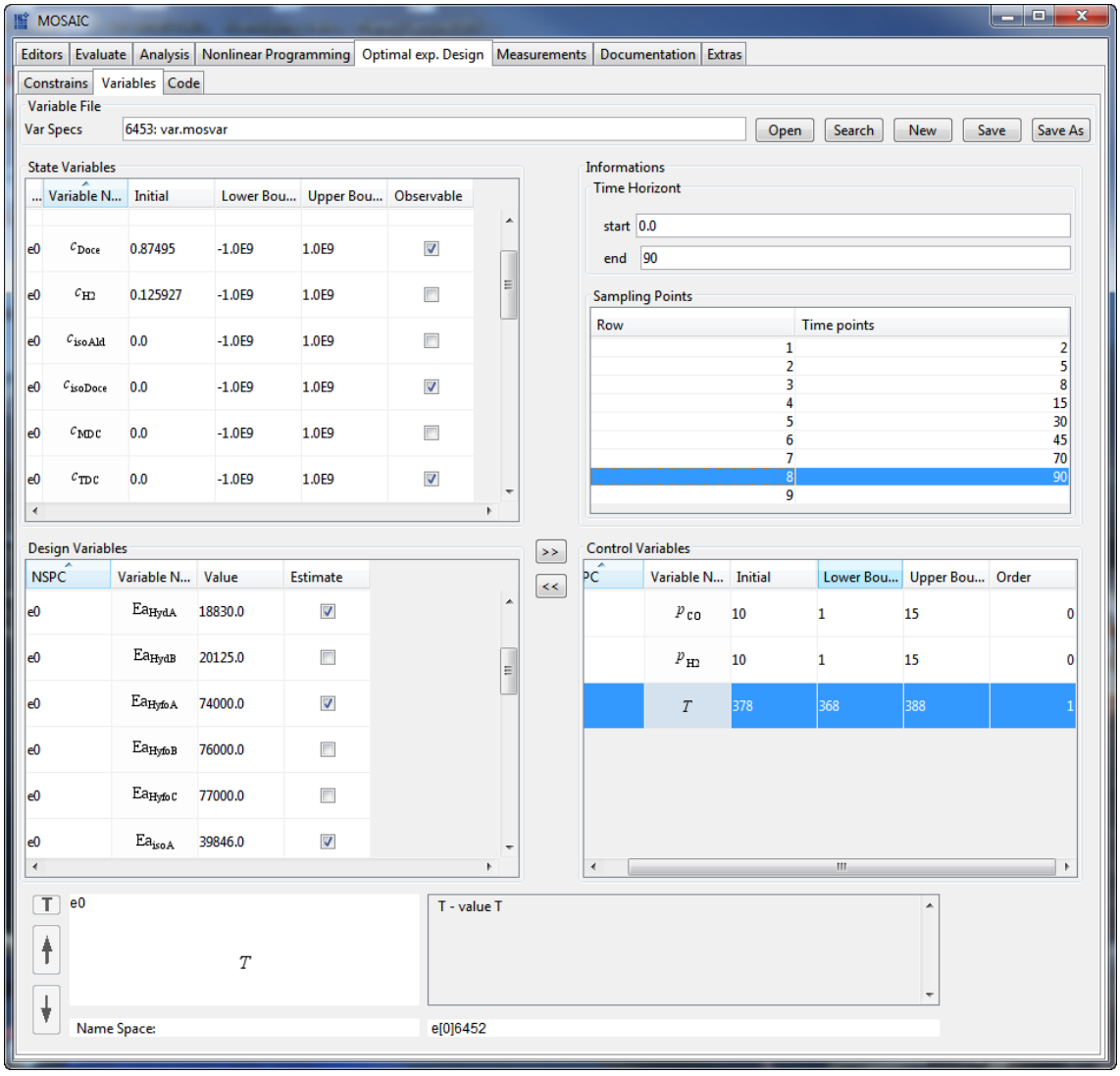


Figure 6.22 MOSAIC – OED interface

7 References

Albersmeyer, J. & Bock, H.G., 2009. *Efficient sensitivity generation for large scale dynamic systems*, Erlangen.

Aldrich, J., 1997. R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3), pp.162–176.

Arellano-Garcia, H., Schöneberger, J. & Körkel, S., 2007. Optimale Versuchsplanung in der chemischen Verfahrenstechnik. *Chemie Ingenieur Technik*, 79(10), pp.1625–1638.

Asprey, S.P. & Macchietto, S., 2002. Designing robust optimal dynamic experiments. *Journal of Process Control*, 12(4), pp.545–556.

Bard, Y., 1973. *Nonlinear parameter estimation*, Academic Press Inc.

Barz, T. et al., 2011. An efficient sparse approach to sensitivity generation for large-scale dynamic optimization. *Computers & Chemical Engineering*, 35, pp.2053–2065.

Barz, T. et al., 2013. Experimental Evaluation of an Approach to Online Redesign of Experiments for Parameter Determination. *AIChE Journal*, 59(6).

Bauer, I. et al., 2000. Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics*, 120, pp.1–25.

Bernas, A. et al., 2010. Kinetics and mass transfer in hydroformylation-bulk or film reaction? *The Canadian Journal of Chemical Engineering*, 88, pp.618–624.

Bhanage, B.M. et al., 1997. Kinetics of hydroformylation of 1 -dodecene using homogeneous HRh (CO) (PPh ,) 3 catalyst. , 169(96), pp.247–257.

Biegler, L.T., 2010. *Nonlinear programmning*, SIAM.

Bock, G.H. & Pitt, J.K., 1984. A Multiple Shooting Algorithm for Direct Solution of Optimal control Problems. In *Ninth IFAC World Congress*. Budapest.

Bonnans, F.J. et al., 2009. *Numerical Optimization: Theoretical and Practical Aspects*, Springer.

Box, G.E.P. & Draper, N.R., 1987. *Empirical Model-Building and Response Surfaces*, John Wiley & Sons Inc.

Brooks, A.N. & Hughes, T.J.R., 1982. Streamline upwind/Petrov-Galerkin Formulations For Convection Dominated Flows With Particular Emphasiss on the Incompressible Navier-Stokes Equations. *Computer Methods in Applied Mechanics and Engineering*, 32, pp.199–259.

Bruce, M., 2013. *McCarl GAMS User Guide*, GAMS Development Corperation.

Brunsch, Y. & Behr, A., 2013. Temperature-controlled catalyst recycling in homogeneous transition-metal catalysis: minimization of catalyst leaching. *Angewandte Chemie*, 52(5), pp.1586–9.

Buzzi-Ferraris, G. & Manenti, F., 2009. Kinetic models analysis. *Chemical Engineering Science*, 64(5), pp.1061–1074.

Carey, G.F. & Finlayson, B.A., 1975. Orthogonal Collocation on Finite Elements. *Chemical Engineering Science*, 30, pp.587–596.

Cervantes, A.M. & Biegler, L.T., 2000. Optimization strategies for dynamic systems. In C. Floudas & P. Pardalos, eds. *Encyclopedia of optimization*. Springer.

Cramér, H., 1946. *Mathematical Methods of Statistics*, Princeton Univ. Press.

Englezos, P. & Kalogerakis, N., 2000. *Applied Parameter Estimation for Chemical Engineers*, Crc Pr Inc.

Espie, D. & Macchietto, S., 1989. The optimal design of dynamic experiments. *AIChE Journal*, 35(2), pp.223–229.

Finlayson, B.A., 1980. *Nonlinear Analysis in Chemical Engineering*, Mcgraw-Hill College.

Fisher, R.A., 1971. *Design of Experiments*, Hafner Press.

Floudas, C. a & Gounaris, C.E., 2008. A review of recent advances in global optimization. *Journal of Global Optimization*, 45(1), pp.3–38.

Floudas, C.A. & Panos, 2009. *Encyclopedia of optimization* second., Springer.

Franceschini, G. & Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63(19), pp.4846–4872.

Franceschini, G. & Macchietto, S., 2007. Validation of a Model for Biodiesel Production through Model-Based Experiment Design. *Industrial & Engineering Chemistry Research*, 46(1), pp.220–232.

Galvanin, F. et al., 2011. Model-Based Design of Experiments in the Presence of Continuous Measurement Systems. *Industrial & Engineering Chemistry Research*, 50(4), p.A–I.

Galvanin, F. et al., 2007. Model-Based Design of Parallel Experiments. *Industrial & Engineering Chemistry Research*, 46, pp.871–882.

Galvanin, F. et al., 2009. Optimal Design of Clinical Tests for the Identification of Physiological Models of Type 1 Diabetes Mellitus. *Industrial & Engineering Chemistry Research*, 48(4), pp.1989–2002.

Geer, S.A. Van De, 2008. Least-Squares Estimation. In *Encyclopedia of Statistics in Quality and Reliability*. John Wiley & Sons, pp. 1041 – 1045.

Ghosh, J.K. & Basu, D., 1988. *Statistical Information and Likelihood*, Springer-Verlag.

Gill, P.E., Murray, W. & Saunders, M.A., 2002. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Journal on Optimization*, 12, pp.979–1006.

Gosset, W.S., 1908. The Probable Error of a Mean. *Biometrika*, 6(1), pp.1–25.

Hamerla, T. et al., 2013. Hydroformylation of 1-Dodecene with Water-Soluble Rhodium Catalysts with Bidentate Ligands in Multiphase Systems. *ChemCatChem*, 5(7), pp.1854–1862.

Haumann, M. et al., 2002. Hydroformylation of 1-dodecene using Rh-TPPTS in a microemulsion. , 225, pp.239–249.

Heidebrecht, P., Sundmacher, K. & Biegler, L.T., 2011. Optimal design of nonlinear temperature programmed reduction experiments. *AIChE Journal*, 57(10), pp.2888–2901.

Hermes, H. & LaSalle, J.P., 1969. *Functional Analysis and Time Optimal Control*, Academic Press.

Hoang, D.M. et al., 2013. Simultaneous Solution Approach to Model-Based Experimental Design. , 00(00), pp.1–15.

Hogg, R. V., McKean, J. & Craig, A.T., 2012. *Introduction to Mathematical Statistics*, Prentice Hall.

Johnson, N.L., Kotz, S. & Balakrishnan, N., 1995. *Continuous Univariate Distributions Volume 2*, John Wiley & Sons.

Kameswaran, S. & Biegler, L.T., 2008. Convergence rates for direct transcription of optimal control problems using collocation at Radau points. *Computational Optimization and Applications*, 41(1), pp.81–126.

Khuri, A.I. & Mukhopadhyay, S., 2010. Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2), pp.128–149.

Kiedorf, G. et al., 2013. Kinetics of 1-dodecene hydroformylation in a thermomorphic solvent system using a rhodium-biphephos catalyst. *Chemical Engineering Science*, pp.1–18.

Kirches, C. et al., 2012. Efficient direct multiple shooting for nonlinear model predictive control on long horizons. *Journal of Process Control of Process Control*, 22, pp.540–550.

Klatt, K.U. & Engell, S., 1993. Rührkesselreaktor mit Parallel- und Folgereaktion. *VDI-Berichte 1026*, pp.101–108.

Koeken, A.C.J. et al., 2011. Full kinetic description of 1-octene hydroformylation in a supercritical medium. *Journal of Molecular Catalysis A: Chemical*, 346(1-2), pp.1–11.

Körkel, S. et al., 2004. Numerical Methods for Optimal Control Problems in Design of Robust Optimal Experiments for Nonlinear Dynamic Processes Parameter estimation and experimental design. *Optimization Methods and Software*, 19, pp.327–338.

Kraume, M., 2013. Integrierte chemische Prozesse in flüssigen Mehrphasensystemen. *Chemie Ingenieur Technik*, 85(10), pp.1499–1511.

Kuntsche, S. et al., 2011. MOSAIC a web-based modeling environment for code generation. *Computers & Chemical Engineering*, 35(11), pp.2257–2273.

Le, L.C., 1990. Maximum Likelihood : An Introduction. *International Statistical Review*, 58(2), pp.153–171.

Leeuwen, P.W. van & Claver, C., 2008. *Rhodium Catalyzed Hydroformylation*, Springer.

Li, S. & Petzold, L., 1999. *Design of new Daspk for Sensitivity Analysis*, Santa Barbara.

Luenberger, D.G., 1984. *Linear and Nonlinear Programming*, Addison Wesley.

Mehra, R.K., 1974. Optimal Inputs for Linear System Identification. *IEEE Transactions on Automatic Control*, 19(3), pp.192–200.

Montgomery, D.C., 2001. *Design and Analysis of Experiments* Fifth., Wiley & Sons.

Müller, D. et al., 2013. Towards a novel process concept for the hydroformylation of higher alkenes: Mini-plant operation strategies via model development and optimal experimental design. *Chemical Engineering Science*, pp.1–12.

Myers, R.H., Montgomery, D.C. & Anderson-Cook, C.M., 2009. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, JOHN WILEY & SONS, INC.

Neumaier, A. et al., 2005. A comparison of complete global optimization solvers. *Mathematical Programming*, 103(2), pp.335–356. Available at: http://www.springerlink.com/index/10.1007/s10107-005-0585-4.

Nocedal, J. & Wright, S., 2006. *Numerical Optimization*, Springer.

Olsen, P.A., Rennie, S.J. & Goel, V., 2012. Efficient Automatic Differentiation of Matrix Functions. *Lectures in Computational Science and Eingineering*, 87, pp.71–81.

Phillips, P.C.B., 1982. The True Characteristic Function of the F Distribution. *Biometrika*, 69(1), pp.261–264.

Pratt, J.W., 1976. F. Y. Edgeworth and R.A. Fisher on the Efficiency for maximum likelihood estimation. *The annals of statistics*, 4(3), pp.501–514.

Rao, C.R., 1945. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37, pp.81–89.

Raol, J.R., Girija, G. & Singh, J., 2004. *Modelling and Parameter Estimation of Dynamic Systems*, Institution of Engineering & Technology.

Rost, A. et al., 2013. Development of a continuous process for the hydroformylation of long-chain olefins in aqueous multiphase systems. *Chemical Engineering and Processing: Process Intensification*, 67, pp.130–135.

Sager, S., 2013. Sampling decisions in optimum experimental design in the light of pontryagin's maximum principle. *SIAM J. on Control and Optimization*, 51(4), pp.3181–3207.

Sahinidis, N. V., 1996. BARON: A general purpose global optimization software package. *Journal of Global Optimization*, 8(2), pp.201–205.

Schöneberger, J. et al., 2009. Model-Based Experimental Analysis of a Fixed Bed Reactor for Catalytic SO 2 Oxidation. *Industrial & Engineering Chemistry Research*, 48, pp.5165–5176.

Senn, S. & Richardson, W., 1994. The first t-test. *Statistics in medicine*, 13(8), pp.785–803.

Sheynin, O., 1995. Helmert's work in the theory of errors. *Archive for History of Exact Sciences*, 49(1), pp.73–104.

Shivakumar, K. & Biegler, L.T., 2006. Simultaneous dynamic optimization strategies: Recent advances and challenges. *Computers and Chemical Engineering*, 30, pp.1560–1575.

Smith, F.L. & Harvey, A.H., 2007. Avoid Common Pitfalls When Using Henry ' s Law. *CEP*, pp.33–39.

Steinberg, D.M. & Hunter, W.G., 1984. Experimental Design : Review and Comment. *TECHNOMETRICS*, 26(2).

Telen, D. et al., 2012. Optimal experiment design for dynamic bioprocesses: A multi-objective approach. *Chemical Engineering Science*, 78, pp.82–97.

Vasantharajan, S. & Biegler, L.T., 1990. Simultaneous Strategies for Parameter Optimization and Accurate Solution of Differential-Algebraic Systems. *Comput. Chem. Eng.*, 14(8), p.1083.

Venables, W.N. & Ripley, B.D., 2003. *Modern Applied Statistics with S*, Springer.

Wong, W.K., 1994. Comparing robust properties of A, D, E and G-optimal designs. *Computational Statistics & Data Analysis*, 18(4), pp.441–448.

Wright, S.J., 1987. *Primal-Dual Interior-Point Methods*,

Zerry, R. et al., 2004. Web-based object oriented modelling and simulation using mathml. *Computer Aided Chemical Engineering*, 18, pp.1171–1176.