

On Decoupling Chemical Reaction Systems - Methods, Analysis and Applications.

THÈSE N° 7376 (2017)

PRÉSENTÉE LE 24 MARS 2017

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE D'AUTOMATIQUE - COMMUN
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Sriniketh SRINIVASAN

acceptée sur proposition du jury:

Dr A. Karimi, président du jury
Prof. D. Bonvin, Dr J. Billeter, directeurs de thèse
Prof. E. Walter, rapporteur
Prof. S. Narasimhan, rapporteur
Dr T. Meyer, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

To my parents and Pavithra

Acknowledgements

This dissertation would not have materialized without the unconditional support of my two advisors, Prof. Dominique Bonvin and Dr. Julien Billeter. Prof. Bonvin, thank you very much for giving me the opportunity to do my PhD at LA. Thanks a lot for your patience, continued optimism and support, as I attempted to find my footing in research. Thank you for guiding me through this beautiful adventure, while at the same time allowing a high degree of independence in research. Your enthusiasm and outlook to life has served as a great source of inspiration for me. Julien, thank you so much for your perspective, comments, and time throughout all the stages of this dissertation.

Besides my supervisors, I am immensely grateful to my predecessor and collaborator Dr. Nirav Bhatt for being a huge source of encouragement and playing a pivotal role in the success of this thesis. I will also take this opportunity to thank Dr. Philippe Müllhaupt, Dr. Christophe Salzmann and Sandra for their immense support, both technical and non-technical, throughout my stay at LA.

Throughout my dissertation, I had opportunities to interact and collaborate with highly competent scientists. Thanks are due to Dr. Kris Villez and Dr. Alma Masic from EAWAG for a successful collaboration and for many interesting discussions. I extend my thanks to my colleague Diogo Rodrigues for stimulating discussions that helped improve numerous results in this dissertation. Thanks also to Darsha Kumar and Srinesh Shenoi for their support.

I warmly thank the president of the defense jury Prof. Alireza Karimi, and the external jury members Prof. Eric Walter, Prof. Shankar Narasimhan, and Dr. Thierry Meyer for reviewing the dissertation and providing many valuable comments.

I would like to thank the other professors at LA, namely, Prof. Roland Longchamp, Prof. Colin Jones, Prof. Alireza Karimi and Prof. Giancarlo Ferrari for maintaining an enjoyable working atmosphere. My sincere thanks to Ruth, Francine, Sara, Eva and Margot for their assistance in administrative matters, and to Francis and Norbert for their assistance in IT.

I would also like to thank all the amazing colleagues I have had an opportunity to interact and work with at LA. I thank my friend Niranjan for being an integral part of this journey. I also thank the members of Cossonay Cricket Club for their continued support. Finally, I thank my family for their support and encouragement throughout.

Abstract

Chemical reaction systems act as the basis to get the desired products from raw materials. An in-depth understanding of all the underlying rate processes is necessary for monitoring, control and optimization of chemical reaction systems. Traditional representation of a reaction system by means of the conservation equations (material and energy balances) leads to a set of highly coupled differential equations. These coupled ODEs provides overall contributions of all the underlying rate processes, and hence, it is difficult to analyse the effect of each rate process in a reaction system. In this thesis, an alternative representation of reaction systems in terms of decoupled variables, namely, vessel extents is reintroduced. The advantages of using the representation in terms of the decoupled variables over the traditional representation are investigated for data reconciliation, model identification and parameter estimation, and state reconstruction and estimation.

Vessel extents of reaction, mass transfer, inlet and initial conditions

In the vessel-extent representation of a chemical reaction system, each extent varies due to a single rate process. For example, the vessel extent of the i th reaction is affected only by the rate of the i th reaction. For a homogeneous reaction system with S species, R reactions, p inlet and one outlet stream, the representation in terms of vessel extents consists of R vessel extents of reaction, p vessel extents of inlet and one extent of initial conditions. In the case where heat effects in a chemical reactor are modeled, an additional extent of heat exchange is introduced in the vessel-extent representation to capture this phenomenon. In this dissertation, we show that there exists a linear transformation that transforms the representation in terms of vessel extents to the traditional representation in terms of material and heat balances and vice versa.

For a multi-phase reaction system, with S_f species in each phase, R_f reactions, p_m mass transfers, p_f inlet streams and an outlet stream, we show that the representation in terms of vessel extents for each phase F is written in terms of R_f vessel extents of reaction, p_m vessel extents of mass transfer, p_f vessel extents of inlet and one extent of initial conditions.

Data reconciliation

The numbers of moles measured during the course of a reaction are typically corrupted by random measurement noise. Data reconciliation techniques improve the accuracy of these measurements by using redundancies in the material and energy balances expressed as relationships between measurements. Since, in the absence of a kinetic model, these relationships cannot integrate information regarding past measurements, they are expressed in the form of algebraic constraints. This dissertation shows that, even in

the absence of a kinetic model, one can use shape constraints to relate measurements at different time instants, thereby improving the accuracy of reconciled estimates. The construction of these shape constraints depends on the operating mode of the chemical reactor. Moreover, this thesis also shows that the representation of the reaction system in terms of vessel extents helps identify additional shape constraints, thereby further improving the quality of the measured numbers of moles. A procedure for deriving shape constraints from measurements is also described.

Sequential kinetic modeling

Modeling chemical reaction systems is an important task. The identified kinetic model must be able to explain all the underlying rate processes such as chemical reactions and mass transfers. Traditionally, the modeling task is carried out using a simultaneous approach, which, for model prediction, requires having a model candidate for all rate processes. The simultaneous approach, which leads to optimal parameter estimates, can be computationally expensive due to its combinatorial nature. The incremental approach, either via rates or extents, was introduced as an alternative to the simultaneous approach. It is characterized by the fact that each rate process is modeled individually, that is, independently of the other rate processes, thus making it computationally more attractive, however at the price of not guaranteeing bias-free estimates. In this dissertation, we propose a novel *sequential* approach that combines the advantages of the incremental and simultaneous approaches.

State estimation

State estimation techniques are used for improving the quality of measured signals and for reconstructing unmeasured quantities during process operation. For chemical reaction systems, nonlinear estimators are often used to improve the quality of estimated concentrations. Usually, these nonlinear estimators, which include the extended Kalman filter, the receding-horizon nonlinear Kalman filter and the moving-horizon estimator, use a state-space representation in terms of numbers of moles. In this dissertation, we formulate the state estimation problem in terms of vessel extents, which allows imposing additional shape constraints in terms of the sign, monotonicity and concavity/convexity properties of extents. We show that the addition of shape constraints often leads to significantly improved state estimates.

Keywords:

Chemical reaction systems; Homogeneous reactions; Multi-phase reaction systems; Vessel extents of reaction; Vessel extents of mass transfer; Kinetic identification; Data reconciliation; State estimation.

Résumé

Les systèmes de réaction chimiques sont à la base des transformations qui permettent d'obtenir des produits désirés à partir de matières premières. Une compréhension approfondie de toutes les vitesses des phénomènes physiques et chimiques mis en jeu est nécessaire pour le monitoring, le contrôle et l'optimisation de ces systèmes de réaction. La représentation traditionnelle d'un système de réaction au moyen d'équations de conservation (conservation de la matière et de l'énergie) mène à un ensemble d'équations différentielles fortement couplées. Ces équations différentielles ordinaires (EDO) représentent l'ensemble des contributions des vitesses de tous les phénomènes impliqués dans un système réactionnel, et cette représentation rend donc difficile l'analyse individuelle de l'effet de chacune de ces vitesses. Dans cette thèse, une représentation alternative des systèmes de réaction se basant sur des variables découplées, appelées *avancements généralisés*, est rappelée. Les avantages de cette représentation en termes de variables découplées par rapport à la représentation traditionnelle sont étudiés dans le contexte de la réconciliation de données, de l'identification des modèles et de l'estimation de leurs paramètres, ainsi que de la reconstruction et de l'estimation d'état.

Avancements généralisés de réaction, de transfert de masse, d'entrées et de conditions initiales

Dans la représentation d'un système de réaction chimique en termes d'avancements généralisés, chaque avancement varie suivant un phénomène physique ou chimique unique. Par exemple, l'avancement ¹ de la i -ème réaction n'est affecté que par la vitesse de la i -ème réaction. Pour un système de réaction homogène avec S espèces chimiques, R réactions, p entrées et une sortie, la représentation en termes d'avancements généralisés se compose de R advancements (généralisés) de réaction, p advancements d'entrée et un avancement de conditions initiales. En cas d'échange de chaleur entre le réacteur et son environnement, un avancement supplémentaire, dit d'échange de chaleur, est introduit dans la représentation en advancements généralisés afin de capturer l'effet de ce phénomène thermique. Dans cette dissertation, nous montrons qu'il existe une transformation linéaire qui transforme la représentation en termes d'avancements généralisés en la représentation traditionnelle exprimées en termes de bilans de matière et d'énergie, et vice versa.

Pour un système réactionnel multiphasique, avec S_f espèces chimiques dans chacune des phases, R_f réactions, p_m transferts de masse, p_f entrées et une sortie, nous montrons que la représentation en termes d'avancements généralisés pour chacune des phases F peut être décrite par R_f advancements (généralisés) de réaction, p_m advancements de transfert de masse, p_f advancements d'entrée et un avancement de conditions initiales.

¹ Pour des raisons de simplicité, le terme *avancement* est parfois employé seul en lieu et place de sa dénomination complète *avancement généralisé*.

Reconciliation de données

Les mesures du nombre de moles de chaque espèce chimique prises au cours d'une réaction sont généralement corrompues par du bruit de mesure aléatoire. Les techniques de reconciliation de données améliorent la précision de ces mesures en utilisant les redondances dans les bilans de matière et d'énergie exprimés sous forme de relations entre les mesures. Ces relations ne pouvant pas intégrer d'information sur les mesures passées en l'absence d'un modèle cinétique, ces équations sont exprimées sous la forme de contraintes algébriques. Cette dissertation montre que, même en l'absence d'un modèle cinétique, des contraintes dites de forme peuvent être employées pour relier les mesures à différents instants, améliorant d'autant la précision des données réconciliées. La relation mathématique de ces contraintes de forme dépend du mode de fonctionnement du réacteur chimique. De plus, cette thèse montre également que la représentation du système réactionnel à l'aide d'avancements permet d'identifier des contraintes de forme supplémentaires, permettant d'améliorer encore la qualité des mesures exprimées en nombres de moles. Une procédure pour dériver ces contraintes de forme à partir de mesures est également décrite.

Identification séquentielle des modèles

L'identification d'un modèle cinétique décrivant un système de réaction chimique est une tâche complexe mais importante. En effet, le modèle cinétique identifié doit être en mesure d'expliquer toutes les vitesses des phénomènes physiques et chimiques mis en jeu, tels que les réactions chimiques et les transferts de masse. Traditionnellement, la tâche de modélisation est effectuée en utilisant une approche simultanée qui nécessite d'avoir un modèle candidat pour la vitesse de chacun des phénomènes. L'approche simultanée conduit à l'estimation optimale des paramètres, mais peut s'avérer coûteuse en temps de calcul en raison de sa nature combinatoire. L'approche incrémentale, basée sur les vitesses ou sur les advancements généralisés, a été présentée comme une alternative à l'approche simultanée. Elle est caractérisée par le fait que la vitesse de chaque phénomène est modélisée individuellement, c'est-à-dire indépendamment des autres vitesses, ce qui la rend plus performante en terme de temps de calcul, au risque cependant de fournir des estimations biaisées. Dans cette dissertation, nous proposons une nouvelle approche, dite séquentielle, qui combine les avantages des approches incrémentales et simultanées.

Estimation d'état

Les techniques d'estimation d'état sont utilisées pour améliorer la qualité des signaux mesurés et pour reconstruire des quantités non mesurées durant le fonctionnement d'un procédé. Pour les systèmes de réaction chimiques, des estimateurs non linéaires sont souvent utilisés pour améliorer la qualité des concentrations estimées. En général, ces estimateurs non linéaires, tels que le filtre de Kalman étendu, le filtre de Kalman non linéaire à horizon glissant, et l'estimateur à horizon glissant, utilisent une représentation d'état en termes de nombre de moles. Dans cette dissertation, nous formulons

le problème d'estimation d'état en termes d'avancement généralisés, permettant ainsi d'imposer des contraintes de forme supplémentaires en termes de signe, de monotonicité, et de concavité / convexité des avancements. Nous montrons que l'ajout de ces contraintes de forme mène souvent à une amélioration significative de l'estimation d'état.

Mots-clés :

Systèmes de réaction chimiques; Réactions homogènes; Systèmes réactionels multiphasiques; Avancements généralisés de réaction; Avancements généralisés de transfert de masse; Identification cinétique; réconciliation de données; Estimation d'état.

Contents

Abstract	iv
Résumé	vi
Nomenclature	xvii
1 Introduction	1
1.1 Motivation	1
1.2 State of the Art	3
1.2.1 Mathematical models	3
1.2.2 Data reconciliation	4
1.2.3 Model identification	5
1.2.4 State estimation	7
1.3 Thesis Outline and Main Contributions	7
2 Material and Energy Balances	13
2.1 Homogeneous Reaction Systems	13
2.1.1 Preliminaries	13
2.1.2 Material balance equations	17
2.1.3 Invariant states	20
2.1.4 Heat balance equations	24

2.2	Heterogeneous Reaction Systems.....	25
2.2.1	Material balance equations	25
2.2.2	Invariant states	26
2.3	Summary	28
3	Vessel Extents Representation	29
3.1	Homogeneous Reaction Systems	29
3.1.1	Two-way representation of chemical reaction systems	30
3.1.2	Three-way representation of chemical reaction systems	31
3.1.3	Vessel extents representation	32
3.1.4	Vessel extents representation with heat balance	37
3.1.5	Use of flowrate information	38
3.2	Heterogeneous Reaction Systems.....	40
3.2.1	Vessel extents representation	40
3.2.2	Use of flowrate information	41
3.3	Simulated Examples	42
3.3.1	Homogeneous reaction system	42
3.3.2	Heterogeneous reaction system	43
3.4	Summary	46
4	Data Reconciliation	49
4.1	Homogeneous Reaction Systems	49
4.1.1	Shape constraints	50
4.2	Heterogeneous Reaction Systems.....	56
4.2.1	Shape constraints	56
4.3	Data Reconciliation	59
4.3.1	Reconciliation using numbers of moles	60
4.3.2	Reconciliation using extents	60

4.3.3	Additional processing	61
4.4	Simulated Examples	62
4.4.1	Homogeneous reaction systems	62
4.4.2	Heterogeneous reaction systems	69
4.5	Summary	71
5	Sequential Model Identification	75
5.1	Homogeneous Reaction Systems	76
5.1.1	Model identification	76
5.1.2	Extent-based Incremental Model Identification (IMI)	77
5.1.3	Extent-based Sequential Model Identification (SMI)	77
5.1.4	Bias due to error propagation and interpolation	80
5.2	Simulated Example	83
5.2.1	Homogeneous reaction system	83
5.3	Summary	90
6	State Reconstruction and Estimation	93
6.1	State Reconstruction	93
6.1.1	State reconstruction in terms of numbers of moles	94
6.1.2	State reconstruction in terms of vessel extents	95
6.2	State Estimation	96
6.2.1	Measurement-based constraints	97
6.2.2	System representation	97
6.2.3	Receding-horizon nonlinear Kalman filter	99
6.3	Simulated Example	101
6.4	Summary	104
7	Conclusions	105

7.1	Summary of Main Results	105
7.2	Outlook and Perspectives	107
A	Data Reconciliation - Heterogeneous Reaction Systems.....	109
A.1	Reconciliation using Numbers of Moles	109
A.2	Reconciliation using Extents	109
B	Sequential Model Identification - Heterogeneous reaction system ...	111
B.1	Extent-based Sequential Model Identification (SMI)	111
C	Reaction Systems with Instantaneous Equilibria.....	113
C.1	Balances for Kinetic Species and Equilibrium Components	114
C.2	Reaction Systems with Equilibrium Reactions	116
C.2.1	Using vessel extents of kinetic reactions	117
C.3	Simulated Example	118
C.4	Summary	123
D	Extent-based Calibration of Spectroscopic Data	125
D.1	Factorization of Spectroscopic Data from Reaction Systems	126
D.1.1	Standard factorization	126
D.1.2	Factorization in vRMV-form.....	127
D.1.3	Factorization in RV-form.....	127
D.2	Prediction of Calibrated Concentrations from Spectroscopic Data	128
D.2.1	Standard calibration.....	128
D.2.2	Calibration in vRMV-form	129
D.2.3	Calibration in vRV-form	130
D.2.4	Choice of calibration model.....	130
D.3	Simulated Example	131
D.4	Summary	134

References	135
-------------------------	------------

Curriculum Vitae	143
-------------------------------	------------

Nomenclature

Accents

$\overset{\circ}{(\cdot)}$	interpolated quantity
$\overset{\circ\circ}{(\cdot)}$	second derivative
$\overset{\cdot}{(\cdot)}$	first derivative
$\overset{\wedge}{(\cdot)}$	reconciled quantity
$\overset{\sim}{(\cdot)}$	measured quantity

Calligraphic symbols

\mathcal{M}	set of all candidate models
\mathcal{I}	set of invariant relationships
\mathcal{K}	set of knowledge-based constraints
\mathcal{E}	set of elements
\mathcal{S}	set of species
\mathcal{S}_c	set of equilibrium components
\mathcal{S}_e	set of equilibrium species
\mathcal{S}_k	set of kinetic species
\mathcal{T}	linear transformation for vessel extent representation
\mathcal{T}_{amr}	linear transformation for three-way representation
\mathcal{T}_{asb}	linear transformation for two-way representation
\mathcal{T}	time window
\mathcal{M}	set of candidate models of a rate process
\mathcal{M}	set of measurement-based constraints

Greek symbols

β	stoichiometric numbers
$\boldsymbol{\beta}$	random vector
$\boldsymbol{\chi}$	concatenated extents of reaction and mass transfer in Appendix B ($R_f + p_m$ -dim.)
δ_m	vector of variables expressing the difference between the extents of mass transfer between phases (p_m -dim.)
$\boldsymbol{\mu}$	concatenated extents of inlet and initial conditions ($p + 1$ -dim.)
ν	stoichiometric coefficient
ω	inverse residence time
$\boldsymbol{\Sigma}$	variance-covariance matrix
$\boldsymbol{\theta}$	vector of parameters
φ	unknown functions
φ_n	unknown functions postulated in terms of numbers of moles
φ_x	unknown functions postulated in terms of vessel extents
ξ	batch extent
ζ	mass-transfer rate
$\boldsymbol{\zeta}$	mass-transfer rate vector (p_m -dim.)

Operators

$(\cdot)^+$	Moore-Penrose inverse of (\cdot)
$(\cdot)^{-1}$	inverse of (\cdot)
$(\cdot)^T$	transpose of (\cdot)
$diag(\cdot)$	diagonalize (\cdot)
$rank(\cdot)$	rank of (\cdot)
$E[\cdot]$	expectation of $[\cdot]$

Roman symbols

\mathbf{A}	atomic matrix ($S \times E$) in Chapter 2, or linearization matrix in Chapter 6
\mathbf{a}	measured absorbance spectrum (W -dim.)
\mathbf{a}_0	initial absorbance spectrum (W -dim.)
\mathbf{A}_{in}	matrix of inlet spectra ($p \times W$)
\mathbf{A}_m	matrix of mass-transfer spectra ($p_m \times W$)
\mathbf{A}_{cal}	matrix of calibration spectra ($H \times W$)

B	concatenated matrix ($S \times d$)
C	linearized measurement equation
c	vector of concentrations (S -dim.)
C_{cal}	matrix of calibration concentrations ($H \times S_c$)
$\check{\mathbf{w}}^k$	vector of weight fractions of the k th inlet stream (S -dim.)
$\check{\mathbf{T}}_{in}$	inlet specific heat vector (p -dim.)
$\check{\mathbf{W}}_{in}$	inlet weight-fraction matrix ($S \times p$)
$\check{q}_{ja,in}$	heat flow via heating/cooling fluid in the jacket
E	transformation of equilibrium species to components in Appendix C ($S_c \times S_e$), or matrix of pure component spectra in Appendix D ($S_a \times W$)
F	calibration matrix ($S_c \times W$)
K	gain matrix
M	projection matrix ($S \times p$)
M_w	pure component molecular weight matrix ($S \times S$)
N	stoichiometric matrix ($R \times S$)
n	vector of number of moles (S -dim.)
n₀	vector of initial numbers of moles (S -dim.)
1	vector of appropriate dimension with all elements being 1
P	invariant matrix ($S \times q$)
Q	variance-covariance matrix for state estimator in Chapter 6
r	vector of reaction rates (R -dim.)
r_v	vector of reaction rates in $\frac{\text{moles}}{\text{time}}$ (R -dim.)
S	projection matrix ($S \times R$)
$\Delta \mathbf{H}_r$	vector of reaction enthalpies (R -dim.)
\mathbf{u}_{in}	vector of inlet mass flowrates (p -dim.)
W	weighting matrix
w	state vector in three-way representation in Chapter 3, or stochastic error variables in Chapter 6
W_{in}	inlet-composition matrix ($S \times p$)
W_{m,f}	mass-transfer matrix for the f phase ($S_f \times p_m$)
x	vector of vessel extents
y	state vector in two-way representation
z	concatenated vector ($(S + 1)$ -dim.)
c_p	specific heat capacity of the reaction mixture
c_{p,ja}	specific heat capacity of the heating/cooling fluid
d	dimensionality of the reaction system

E	number of elements
H	final sampling instant
m	mass of reaction mixture
p	number of independent inlets
p_m	number of mass transfers
q	number of invariants
Q_r	heat in the reactor
q_{ex}	heat flow between reaction mixture and jacket
Q_{ja}	heat in the jacket
R	number of independent reactions
S	number of species
T_r	temperature in the reactor
u_{out}	outlet mass flowrate
V	volume inside the reactor
W	number of wavelenghts/wavenumbers

Subscripts

δ_h	variable expressing the difference between the extents of heat exchange in the reactor and jacket
a	related to available species, or absorbing species in Appendix D
c	related to equilibrium component in Appendix C, or related to calibrated species in Appendix D
e	related to instantaneous equilibria in Appendix C
f	related to f phase, $f \in \{g, l\}$
g	related to gas phase
h	related to heat transfer
ic	related to initial conditions
in	related to inlets
iv	related to invariants
k	related to kinetically controlled in Appendix C
l	related to liquid phase
m	related to mass transfer
nc	related to non-removed non-consumed species
np	related to non-added non-produced species
out	related to outlet flow

r related to reactions

Superscripts

MV mass-transfer variant form

RMV reaction- and mass-transfer variant form

RV reaction-variant form

vMV vessel mass-transfer variant form

$vRMV$ vessel reaction- and mass-transfer variant form

vRV vessel reaction-variant form

Chapter 1

Introduction

1.1 Motivation

Chemical reaction systems are widely used by various industries to convert raw materials into high-value and marketable products. Some of these chemical reactions are complex and arise in a number of different processes such as combustion, fluid catalytic cracking, chemical-vapour decomposition and in biotechnology. For example, the (bio)chemical industry uses reaction processes to convert raw materials into desired products that include polymers, organic chemicals, vitamins, vaccines and drugs.

In the recent years, fast-changing market conditions have forced the process industry to operate its chemical production facilities in an increasingly transient regime, featuring fast and frequent changes in product type, product grade and production rate [9, 68]. While profit is high on the agenda of these industries, the products must be of very high quality and, at the same time, satisfy the environmental restrictions. Industries try to meet these objectives by building reliable mathematical process models during the process development phase to describe the chemical reaction systems as shown in Fig. 1.1. These process models can considerably reduce process development costs and can be used for better process understanding, simulation and design.

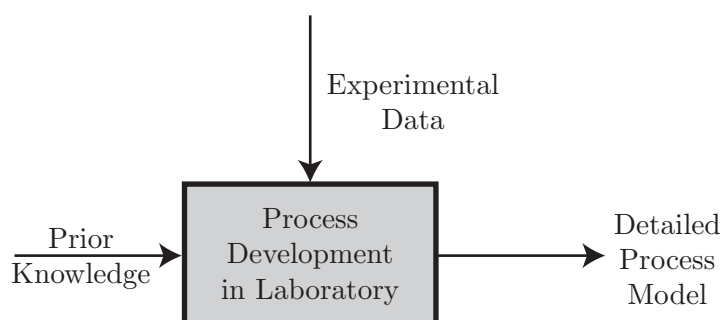


Figure 1.1 Schematic diagram illustrating the modeling phase during process development in the laboratory.

Process models of a chemical reaction system should be able to explain the effects of all the underlying chemical reactions. Apart from reactions, these models must also explain (i) material exchange via inlet/outlet flows, mass transfers, and (ii) energy exchanges

via heating and cooling (heat transfers). The process models could either be derived from first principles [59, 67] or can be empirical in nature such as response surface models [19]. The developed process models can then be used during process operations for model-based monitoring, control and optimization [30, 44, 56] and hence lead to improved safety, productivity and product quality.

Identifying a reliable description of reaction kinetics and transport phenomena represents the main challenge in building process models for chemical reaction systems. In practice, such a description is constructed from experimental data collected in the laboratory as shown in Fig 1.1 [48].

During the process development phase in the laboratory, the measurements such as temperature, pressure and concentrations can be obtained from different sources such as calorimetry, spectrometry (mid-infrared, near-infrared and ultraviolet/visible - UV/VIS) or chromatography. In an ideal world, these measurements taken during the course of a chemical reaction system will give the exact values of the states inside the reactor and can be used for either kinetic modeling or for data-driven approaches. However, in reality, these measurements taken are corrupted by random measurement noise. Since measurements are corrupted by noise, the performance of the modeling/identification task, and thus also of the subsequent monitoring, control and optimization steps, depends highly on the accuracy of the measurements.

For model-based monitoring, control and optimization, the kinetic model identified at the process development stage is used along with the process data from the production environment in order to adjust the process model and to improve the state estimates as shown in Fig. 1.2.

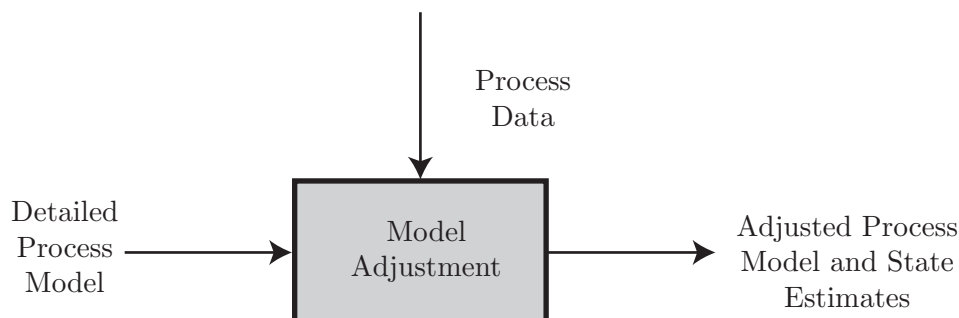


Figure 1.2 Schematic diagram illustrating model adjustment and state estimation during process operations.

In this dissertation, we aim to solve the following problems to meet industrial goals based on experimental data:

- Find a suitable system representation for chemical reaction systems,

- Find a pretreatment for the experimental data in the absence of a process model,
- Develop a methodology for model identification of chemical reaction systems, and
- Develop a procedure for state estimation and model adjustment during process operation.

1.2 State of the Art

1.2.1 Mathematical models

Models of chemical reaction processes based on first principles describe the evolution of states (mass, concentrations, temperature) by means of balance equations of differential nature (continuity equation, molar balances, heat balances) and constitutive equations of algebraic nature (e.g. equilibrium relationships, rate expressions). These models usually include information regarding the underlying reactions (e.g. stoichiometries, reaction kinetics, heats of reaction), the transfers of mass within and between phases, and the operating mode of the reactor (e.g. initial conditions, external exchange terms, operating constraints). The presence of all these phenomena, and in particular their interactions, complicates the analysis and operation of chemical reactors.

To simplify the analysis of chemical reaction systems, researchers have introduced alternative representations of these systems, where each state variable varies due to a single rate process [4, 8]. Ideally, one would like to have true *variants*, whereby each state depends only on one phenomenon, for example a reaction variant that varies only due to the effect of a single reaction. Asbjørnsen and co-workers [7, 8, 33] introduced the concepts of reaction variants and reaction invariants. However, the reaction variants proposed by Asbjørnsen encompass more than the reaction contributions, since they are also affected by the inlet and outlet flows. Hence, Friedly [34, 35] proposed to compute the extents of “equivalent batch reactions”, associating the remainder to transport processes. For open homogeneous reaction systems, Srinivasan et al. [70] developed a *nonlinear* transformation of the numbers of moles to reaction variants, flow variants, and reaction and flow invariants, thereby separating the effects of reactions and flows. Later, Amrhein et al. [4] refined that transformation to make it linear (at the price of losing the one-to-one property) and therefore more easily interpretable and applicable. They also showed that, for a reactor with an outlet flow, the concept of *vessel extent* was most useful, as it represents the amount of material associated with a given process (reaction, exchange) that is still in the reactor. Bhatt et al. [14] extended that concept to heterogeneous gas-liquid reaction systems for the case of no reaction and no accumulation in the film, the result being vessel extents of reaction, mass transfer, inlet and outlet, as well as true invariants that are identically equal to zero.

Various implications of reaction variants/invariants have been studied in the literature. For example, Srinivasan et al. [70] discussed the implications of reaction and flow

variants/invariants for control-related tasks such as model reduction, state accessibility, state reconstruction and feedback linearizability. On the one hand, control laws using reaction variants have been proposed for continuous stirred-tank reactors [40, 78, 29, 32]. The concept of batch extent of reaction is very useful to describe the dynamic behavior of a chemical reaction since a reaction rate is simply the derivative of the corresponding extent of reaction. Bonvin and Rippin [18] used batch extents of reaction to identify stoichiometric models without the knowledge of reaction kinetics. Reaction extents have been used extensively for the kinetic identification of both homogeneous and heterogeneous reaction systems using either concentration [16] or spectroscopic [17] measurements.

On the other hand, the fact that reaction invariants are independent of the reaction progress has also been exploited for process analysis, design and control. For example, reaction invariants have been used to study the state controllability and observability of continuous stirred-tank reactors [12, 33]. Reaction invariants have also been used to automate the task of formulating mole balance equations for the non-reacting part (such as mixing and splitting operations) of complex processes, thereby helping determine the number of degrees of freedom for process synthesis [36]. Furthermore, Waller and Mäkilä [78] demonstrated the use of reaction invariants to control pH, assuming that the equilibrium reactions are very fast. Grüner et al. [39] showed that, through the use of reaction invariants, the dynamic behavior of reaction-separation processes with fast (equilibrium) reactions resembles the dynamic behavior of corresponding non-reactive systems in a reduced set of transformed variables. Aggarwal et al. [1] considered multi-phase reactors operating at thermodynamic equilibrium and were able to use the concept of reaction invariants, which they labeled invariant inventories, to reduce the order of the dynamic model and use it for control.

1.2.2 Data reconciliation

The accuracy and precision of measurements can be improved using data reconciliation (DR) techniques [53, 66] as shown in Fig. 1.3. Static DR uses redundancies in the conservation equations expressed as relationships between measurements to improve the exactness of the measurements [26]. In the case of chemical reaction systems, these relationships are derived from the conservation equations, namely, the material and energy balance equations [27].

However, the algebraic constraints used for DR are valid at each given time instant, that is, not over neighboring time instants and do not relate measurements taken at different times [73]. In such a case, DR solves a weighted least-squares optimization problem at each time instant. If a kinetic model is available, DR can be performed using techniques such as Recursive Nonlinear Dynamic Data Reconciliation (RNDDR) or Moving Horizon Estimator (MHE) [57, 58], that is, data over several time instants can be reconciled together.

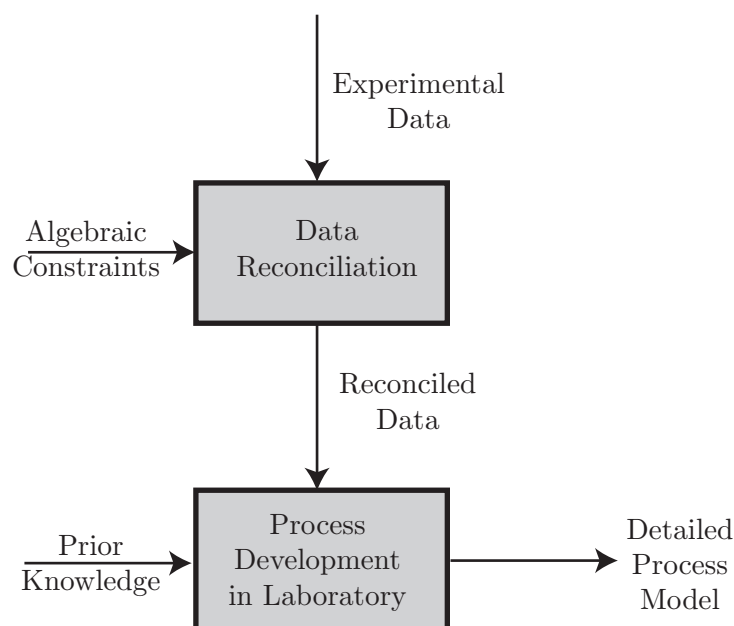


Figure 1.3 Schematic diagram illustrating the integration of data reconciliation in the modeling phase during the process development stage in the laboratory.

1.2.3 Model identification

Model identification of chemical reaction systems involves determining model structures and estimating the corresponding model parameters for all the rate processes from experimental or reconciled data. The model identification task can be carried out either using a simultaneous approach or via an incremental approach.

Simultaneous model identification

Traditionally, model identification is carried out simultaneously for all the rate processes involved in the system, and thus the identification approach is labeled “simultaneous method of parameter estimation”. In this method, a model candidate is proposed for each reaction and mass transfer. The corresponding parameters are estimated by comparing the measured and predicted numbers of moles. The procedure is repeated for all the possible combinations of candidate rate-law models. The best combination is chosen based on a pre-defined criterion, such as a least-squares value, Akaike information criterion, or a Bayesian information criterion. The advantage of simultaneous identification lies in the fact that the model parameters are estimated in a maximum-likelihood sense [10]. However, simultaneous identification suffers from high computational cost when, for each rate process, several model candidates are to be tested. If M models have to be tested for each of the N rate processes, simultaneous identification requires evaluating M^N model combinations. Additionally, since multiple parameters are estimated at once, simultaneous identification might lead to convergence problems [21] and high cor-

relation between the estimated parameters. The convergence issue is generally tackled by solving the regression problem with different initial guesses, while proper design of experiments is required to reduce the correlation between parameters [52].

Incremental model identification

Another modeling perspective is the incremental model identification approach [11, 16, 22, 51]. In contrast to simultaneous identification, the incremental approach deals with each rate process individually, that is, independently of the other rate processes. This is possible by transforming the measured state variables to so-called variant states, either rates or extents, each one depending on a single rate process. Since the modeling of each rate process can be done independently of the other rates, the incremental approach does not have the combinatorial complexity of the simultaneous identification and is therefore less computationally expensive. In addition, the incremental approach is less prone to convergence issues since the dimensionality of each regression problem is significantly reduced compared to the simultaneous identification. On the other hand, the parameters obtained by the incremental approach may be biased because (noisy) *measurements* must be included in the rate expressions as substitutes for the concentrations that cannot be predicted solely by that particular rate process. Furthermore, in the case of the extent-based approach, the measurements must also be interpolated so that they become available at all times for the purpose of integrating the rate laws. The noise present in these measurements typically leads to biased parameter estimates. As a simultaneous identification step is eventually made to obtain maximum-likelihood estimates based on the model structures identified incrementally, the number of regression problems to be solved with the incremental approach is $M \cdot N + 1$. The incremental approach can be of two types, namely, rate-based and extent-based:

- Rate-based incremental identification: In the rate-based approach, the reaction rates are obtained by differentiation of the measured numbers of moles and the knowledge of the stoichiometry [23]. Candidate rate laws are then postulated for each of the rates independently, and their corresponding parameters are estimated by comparing the measured and estimated rates. The rate-based incremental identification is computationally the cheapest kinetic modeling technique. However, the differentiation of noisy measurements can introduce strong biases so that efficient regularization techniques are necessary [21].
- Extent-based incremental identification: In the extent-based approach, on the other hand, the measured numbers of moles are transformed to so-called *vessel extents* [4, 14]. A vessel extent represents a generalization of the concept of batch extent in the presence of an outlet flow. Model identification for each rate process is carried out independently of the other rates by solving an optimization problem, in which the measured and modeled extents are compared. Note that *measured extents* denote extents that have been obtained by transformation of measured numbers of moles and/or indirect measurements thereof such as calorimetry [72] or spectroscopy [17]. The extent-based identification approach involves numerical integration, which

makes it computationally more intensive than the rate-based approach. Nevertheless, the extent-based approach has been shown to yield tighter confidence intervals and to have a better model discrimination capability than the rate-based approach [16]. The main drawback of the extent-based over the rate-based approach lies in the need to interpolate measurements for the integration procedure. Also, the extent-based approach introduced in [15] does not take into account the correlation that exists between the extents.

1.2.4 State estimation

The field of state estimation focuses on both improving the accuracy of the measured signals and reconstructing unmeasured signals by enforcing their consistency with the given process model [69]. Several state estimators are available for nonlinear dynamic systems. Among these estimators, the most commonly used is probably the extended Kalman filter (EKF) [42]. EKF is recursive by nature and thus can easily be implemented in real time. The major drawback of EKF lies in its inability to handle bounds and algebraic constraints, which are common in the representation of chemical reaction systems.

As an alternative approach, recursive nonlinear dynamic data reconciliation (RNDDR) was introduced [75]. This method has the advantage of preserving the prediction step of the EKF method, but the update step is formulated as a constrained optimization problem. In the unconstrained case, this method reduces to the traditional EKF. The moving-horizon estimator (MHE) constitutes an alternative that can handle constraints on the estimated states [57, 58]. A constrained optimization problem is formulated at each sampling time using a time window of past measurements. This allows incorporating shape constraints (such as sign, monotonicity and curvature) in the estimation problem for the given window. The drawback of the MHE method is the need to solve differential equations within the optimization loop, which can become a computational issue for real-time estimation. The receding-horizon nonlinear Kalman filter (RNK) is another nonlinear state estimator. It is based on the prediction and update steps of the Kalman filter [61]. In the update step, an optimization problem is solved using a time window of past measurements.

1.3 Thesis Outline and Main Contributions

Chapter 2: Material and Energy balances We start by reviewing the traditional representation of chemical reaction systems in terms of material and energy balance equations for homogeneous and heterogeneous reaction systems. In the first part of Chapter 2, some basic concepts such as stoichiometry, atomic matrix, independent reactions and independent inlets are introduced. This leads to an important relationship between the number of species and the number of independent reactions in a reaction system. This relationship will be used repeatedly in the following chapters for the

purpose of transforming measured numbers of moles to vessel extents. In the second part of the chapter, the material and energy balance equations for a homogeneous reaction system in a generic open reactor with inlet and outlet streams are discussed. In this part, we introduce the concept of invariant relationships and give a procedure for deriving such relationships from material and energy balances. In the last part of the chapter, the balance equations and invariant relationships are extended to heterogeneous reaction systems with steady-state mass transfer.

Main Contributions

- This chapter introduces a systematic procedure for deriving invariant relationships for both homogeneous and heterogeneous reaction systems.

Chapter 3: Alternative Representation - Vessel Extents This chapter introduces alternative representations of chemical reaction systems in terms of variant and invariant states. First, we briefly review the various representations in terms of variant states introduced in the literature. Next, definitions of state variables called vessel extents, which are functions of a single rate process, are recalled. This leads to an alternative representation for homogeneous reaction systems in terms of vessel extents. The procedure for transforming the representation from numbers of moles to vessel extents and vice versa is explained in detail. The representation in terms of vessel extents is then extended to reactor models with heat balance. In this case, an additional vessel extent of heat exchange is needed to represent the reaction system. In the final part, the representation in terms of vessel extents is extended to heterogeneous reaction systems.

Main Contributions

- A new transformation is introduced that brings the representation in terms of numbers of moles to a representation in terms of the vessel extents.
- Extensions of the vessel-extent representation to reaction systems with a heat balance and to heterogeneous reaction systems are discussed.

Chapter 4: Data Reconciliation This chapter shows that, even in the absence of a kinetic model, dynamic information can be added to the data reconciliation problem to further improve the quality of the reconciled estimates. Shape constraints are introduced to relate measurements at different time instants, thereby improving the accuracy of reconciled estimates. A procedure for constructing the knowledge-based shape constraints in the representation in terms of both the numbers of moles and extents is introduced. A procedure for deriving shape constraints from measurements is also described. Fig. 1.4 illustrates the use of shape constraints during data reconciliation. The performance of the data reconciliation procedure in terms of both the numbers of moles and vessel extents is compared via simulated examples of homogeneous and heterogeneous reaction systems.

Main Contributions

- The concept of shape constraints is introduced to improve the accuracy of data reconciliation procedures.
- Knowledge-based constraints are derived for homogeneous and heterogeneous reaction systems in terms of numbers of moles and in terms of vessel extents. We show that the vessel extent representation of a reaction system often leads to additional shape constraints.
- For cases where knowledge-based constraints do not exist, a procedure for deriving measurement-based constraints is introduced.

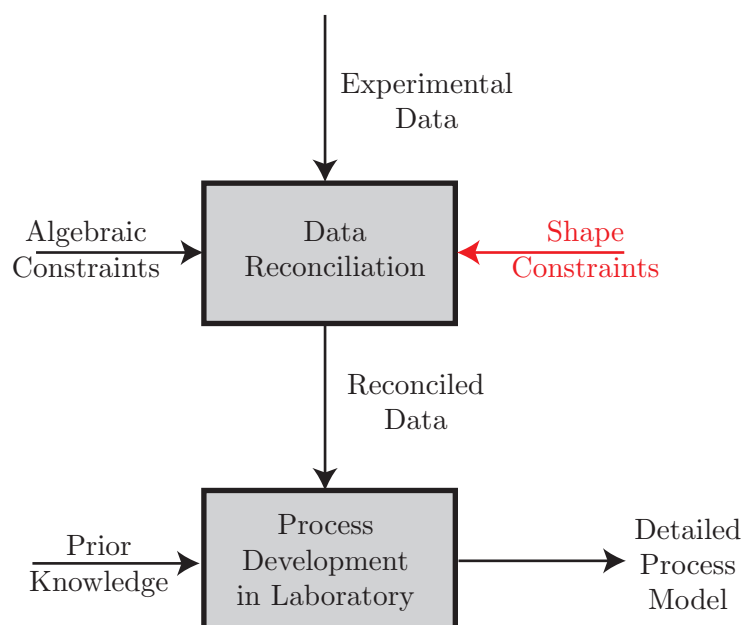


Figure 1.4 Schematic diagram illustrating the integration of shape constraints in the data reconciliation step.

Chapter 5: Sequential Model Identification In this chapter, the reconciled numbers of moles or vessel extents from Chapter 4 are used to identify the kinetics of the reaction system. A novel sequential model identification approach is proposed, which is based on the extent-based incremental approach that combines the advantages of the incremental and simultaneous approaches. In addition, a method to reduce the structural bias that is present in the extent-based incremental approach is introduced. The sequential model identification procedure is illustrated via a simulated example of a homogeneous reaction system.

Main Contributions

- A sequential approach for kinetic modeling of chemical reaction system using the extent-based incremental approach is developed.
- A procedure for removing the structural bias in the incremental approach is discussed.

Chapter 6: State Reconstruction and Estimation In the first part of this chapter, a methodology for reconstructing the numbers of moles of the unmeasured species without the knowledge of a kinetic model is introduced. If a kinetic model for the reaction system is available, state estimation techniques can then be used to improve the quality of measured signals, estimate unmeasured quantities and at the same time to update the model parameters. In this chapter, a nonlinear state estimation problem is formulated by using the knowledge-based shape constraints derived in Chapter 4 as shown in Fig. 1.5. The addition of shape constraints in the state estimation problem leads to significantly improved state estimates. Additionally, the formulation of the state estimation problem in terms of vessel extents performs better than the formulation in terms of numbers of moles. A procedure for deriving the measurement-based constraints is also introduced. The state estimation procedure is illustrated via a simulated example.

Main Contributions

- State reconstruction procedures are introduced for reconstructing the numbers of moles of unmeasured species in the absence of a kinetic model.
- State estimators for chemical reaction systems in the presence of shape constraints are introduced.
- The formulation in terms of vessel extents leads to better state estimates than the formulation in terms of numbers of moles due to additional shape constraints.
- A procedure for identifying measurement-based constraints using the identified model and measurements is explained.

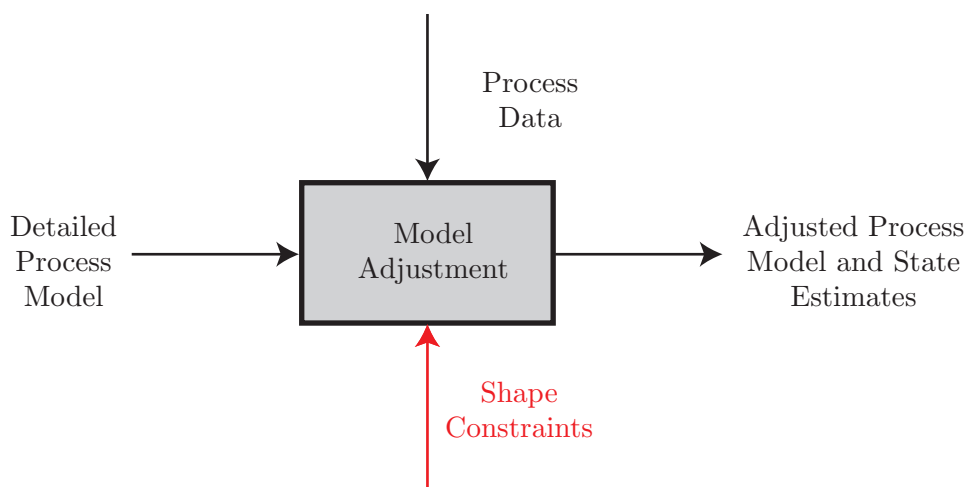


Figure 1.5 Schematic diagram illustrating the integration of shape constraints to the state estimation problem.

Chapter 7: Conclusion Finally, Chapter 7 summarizes the main contributions of this thesis and its results, and discusses some future perspectives.

For pedagogical reasons, the data reconciliation and sequential model identification procedure for heterogeneous reaction systems are presented in Appendix A and B. Appendix C extends the vessel extent representation of a chemical reaction system to reaction systems with instantaneous equilibria. In Appendix D, we discuss a method for using the vessel extents representation for building calibration models of spectroscopic data.

Chapter 2

Material and Energy Balances

This chapter introduces the material and energy balance equations of homogeneous and heterogeneous chemical reaction systems. Section 2.1.1 introduces some basic definitions and concepts required for writing the first-principles models of homogeneous reaction systems. Section 2.1.2 derives the mole balance equations for a homogeneous reaction system in a generic open reactor, a semi-batch reactor and a batch reactor. The concept of system invariants and a procedure for deriving these invariant relationships from material balance equations are proposed in Section 2.1.3. The generic heat balance equations for a reactor and a heating/cooling jacket are derived in Section 2.1.4 along with the combined material and energy balance equations. The invariant relationships are also derived for the combined material and energy balance equations.

Section 2.2 extends the material balance equations introduced for a homogeneous reaction system to a multi-phase heterogeneous reaction system. The mole balance equations are written for a fluid-fluid reaction system with steady-state mass transfer. The system invariant relationships for a heterogeneous reaction system are also reported.

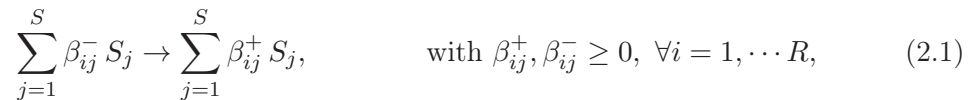
2.1 Homogeneous Reaction Systems

2.1.1 Preliminaries

Consider a homogeneous chemical system with S species living in the set \mathcal{S} , composed of E chemical elements living in the set \mathcal{E} . Let α_{se} denote the number of elements of type e present in species s , then the atomic matrix \mathbf{A} of dimension $S \times E$ can be constructed as:

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1E} \\ \vdots & \vdots & \vdots \\ \alpha_{S1} & \dots & \alpha_{SE} \end{bmatrix}$$

Let the system consist of R chemical reactions represented via the following reaction network:



where β_{ij}^+ and β_{ij}^- are the stoichiometric numbers for the j th species involved in the i th reaction. The net amount of the j th species involved in i th reaction can be described by its stoichiometric coefficient ν_{ij} , defined as:

$$\nu_{ij} = \beta_{ij}^+ - \beta_{ij}^- \quad (2.2)$$

The stoichiometric matrix \mathbf{N} of dimension $R \times S$ containing the stoichiometric coefficients of all the R reactions is defined as:

$$\mathbf{N} = \begin{bmatrix} \nu_{11} & \dots & \nu_{1S} \\ \vdots & \vdots & \vdots \\ \nu_{R1} & \dots & \nu_{RS} \end{bmatrix}$$

The concepts of atomic and stoichiometric matrices are illustrated in Example 2.1.

Reversible reactions: In some cases, the products are reconverted to the reactants, i.e. the reverse reaction is significant as well. The reaction network for reversible reactions are denoted:

$$\sum_{j=1}^S \beta_{ij}^- S_j \rightleftharpoons \sum_{j=1}^S \beta_{ij}^+ S_j \quad \text{where } \beta_{ij}^+, \beta_{ij}^- \geq 0. \quad (2.3)$$

In such cases, in order to maintain consistency with the description of irreversible reactions, the reversible reactions must be encoded as two irreversible reactions [28].

$$\begin{aligned} \sum_{j=1}^S \beta_{ij}^- S_j &\rightarrow \sum_{j=1}^S \beta_{ij}^+ S_j & \text{where } \beta_{ij}^+, \beta_{ij}^- \geq 0 \\ \sum_{j=1}^S \beta_{ij}^+ S_j &\rightarrow \sum_{j=1}^S \beta_{ij}^- S_j \end{aligned}$$

Therefore, the reversible reaction is represented by two rows in the stoichiometric matrix \mathbf{N} , leading to a linearly dependent stoichiometric matrix, that is, $\text{rank}(\mathbf{N}) < R$. Amrhein [2] has developed a procedure for converting the dependent stoichiometric matrix into a linearly independent stoichiometric matrix based on the concept of independent reactions, and net forward reactions.

Example 2.1 (Hydrolysis of Sucrose)

Consider the hydrolysis of sucrose to glucose and fructose,



For this reaction system, the set of elements is $\mathcal{E} = \{C, H, O\}$, while the set of species is $\mathcal{S} = \{C_{12}H_{22}O_{11}, H_2O, C_6H_{12}O_6\}$. The atomic matrix \mathbf{A} is then given by:

$$\mathbf{A} = \begin{bmatrix} 12 & 22 & 11 \\ 0 & 2 & 1 \\ 6 & 12 & 6 \end{bmatrix}$$

The stoichiometric numbers of the three species are $\beta_1^- = \{1, 1, 0\}$ and $\beta_1^+ = \{0, 0, 2\}$. The stoichiometric coefficients of the three species involved in this reaction can be written as:

$$\nu_{C_{12}H_{22}O_{11}} = 0 - 1 = -1$$

$$\nu_{H_2O} = 0 - 1 = -1$$

$$\nu_{C_6H_{12}O_6} = 2 - 0 = 2$$

The stoichiometric matrix \mathbf{N} can then be written as:

$$\mathbf{N} = \begin{bmatrix} -1 & -1 & 2 \end{bmatrix}$$

Definition 2.1 (Independent reactions)

The R reactions are said to be independent if (i) the rows of \mathbf{N} are linearly independent, i.e., $\text{rank}(\mathbf{N}) = R$, and (ii) there exists some finite time interval for which the reaction rate profiles $\mathbf{r}(t)$ are linearly independent, i.e., $\beta^T \mathbf{r}(t) = 0, \forall t \in [t_0, t_1] \Leftrightarrow \beta = \mathbf{0}_R$.

The concept of independent reactions is illustrated in Example 2.2.

Example 2.2 (Ethanolysis of Phthalyl Chloride)

In this reaction system, phthalyl chloride (A) and ethanol (B) react to form phthalyl chloride monoethyl ester (C), phthalic diethyl ester (E) and hydrochloric acid (D). Ethanol and hydrochloric acid also react reversibly to form ethyl chloride (F) and water (G):



For this reaction system, the stoichiometric matrix \mathbf{N} based on [2] is written as:

$$\mathbf{N} = \begin{bmatrix} -1 & -1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 1 \end{bmatrix}$$

Throughout this thesis, and without loss of generality, all the R reactions will be assumed to be independent.

Relationship between number of species and number of reactions

For a given reaction system with S species and R independent reactions, chemical elements are transferred from one species to another but are conserved [5],

$$\mathbf{N}\mathbf{A} = \mathbf{0}_{R \times E}. \quad (2.4)$$

Based on the rank-nullity theorem [74],

$$\text{rank}(\mathbf{N}) + \text{rank}(\mathbf{A}) = S,$$

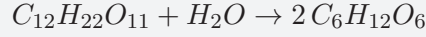
which leads to the following relationship for the number of independent reactions for a given set of species \mathcal{S} :

$$R = \text{rank}(\mathbf{N}) = S - \text{rank}(\mathbf{A}). \quad (2.5)$$

Since $\text{rank}(\mathbf{A}) \geq 1$, the maximum number of independent reactions possible in a chemical reaction system, R_{max} , is always strictly less than the number of species, that is, $R_{max} < S$.

Example 2.1 (Hydrolysis of Sucrose cont'd..)

Consider the hydrolysis of sucrose to glucose and fructose,



This reaction has three species $\mathcal{S} = \{C_{12}H_{22}O_{11}, H_2O, C_6H_{12}O_6\}$ and the atomic matrix:

$$\mathbf{A} = \begin{bmatrix} 12 & 22 & 11 \\ 0 & 2 & 1 \\ 6 & 12 & 6 \end{bmatrix}$$

The rank of the atomic matrix \mathbf{A} is 2. This indicates that this reaction has at most $S - \text{rank}(\mathbf{A}) = 3 - 2 = 1$ independent reaction.

Inlet streams

Let the reaction system consists of p inlet streams that are fed into the reaction system with mass flowrates \mathbf{u}_{in} . The matrix \mathbf{W}_{in} of dimension $S \times p$ is the inlet composition matrix.

Definition 2.2 (Independent inlets)

The p inlet streams are said to be independent if (i) the columns of \mathbf{W}_{in} are linearly independent, i.e., $\text{rank}(\mathbf{W}_{in}) = p$, and (ii) there exists some finite time interval for which the inlet mass flowrate profiles $\mathbf{u}_{in}(t)$ are linearly independent, i.e., $\beta^T \mathbf{u}_{in}(t) = 0$, $\forall t \in [t_0, t_1] \Leftrightarrow \beta = \mathbf{0}_p$.

Throughout this thesis, and without loss of generality, it is assumed that the p inlet streams are independent.

2.1.2 Material balance equations

The mole balance equations for a homogeneous reaction system involving S species, R reactions, p inlet streams, and one outlet stream can be written as follows:

$$\dot{\mathbf{n}}(t) = \mathbf{N}^T \mathbf{r}_v(t) + \mathbf{W}_{in} \mathbf{u}_{in}(t) - \omega(t) \mathbf{n}(t), \quad \mathbf{n}(0) = \mathbf{n}_0, \quad (2.6)$$

where $\mathbf{n}(t)$ is the S -dimensional vector of numbers of moles, \mathbf{N} denotes the $R \times S$ stoichiometric matrix, $\mathbf{r}_v(t) := V(t) \mathbf{r}(t)$ is the R -dimensional vector of reaction rates expressed in $\frac{\text{moles}}{\text{time}}$ and \mathbf{r} the corresponding vector of reaction rates, with V the volume, $\mathbf{W}_{in} = \mathbf{M}_w^{-1} \check{\mathbf{W}}_{in}$ is the $S \times p$ matrix of inlet compositions, with \mathbf{M}_w the S -dimensional diagonal matrix of molecular weights and $\check{\mathbf{W}}_{in} = [\check{\mathbf{w}}_{in}^1 \cdots \check{\mathbf{w}}_{in}^p]$ with $\check{\mathbf{w}}_{in}^j$ the S -dimensional vector of weight fractions of the j th inlet flow, \mathbf{u}_{in} is the p -dimensional

mass flowrate, $\omega := \frac{u_{out}(t)}{m(t)}$ is the inverse residence time with the mass m and the outlet mass flowrate u_{out} , and \mathbf{n}_0 is the S -dimensional vector of initial conditions. Note that, since the reaction rates are modeled as *unknown time signals* $\mathbf{r}_v(t)$, the mole balance equations (2.6) hold independently of the operating conditions. The reaction rates $\mathbf{r}_v(t)$ should formally be formulated as functions of the numbers of moles, namely, $\mathbf{r}_v(t) := V(t) \boldsymbol{\varphi}_n(\mathbf{n}(t), \boldsymbol{\theta})$. The task of model identification consists in identifying these functions and their corresponding parameters $\boldsymbol{\theta}$ from experimental data. If necessary, the concentrations $\mathbf{c}(t)$ can be constructed from the numbers of moles using the relationship $\mathbf{c}(t) = \frac{\mathbf{n}(t)}{V(t)}$.

The flowrates $\mathbf{u}_{in}(t)$ and $u_{out}(t)$ are considered as independent (input) variables in Eq. (2.6). The way these variables are adjusted depends on the particular experimental situation, for example, some elements of \mathbf{u}_{in} are adjusted to control the temperature in a semi-batch reactor, or u_{out} is a function of the inlet flows in a constant-mass reactor. The continuity equation (or total mass balance) is given by:

$$\dot{m}(t) = \mathbf{1}_p^T \mathbf{u}_{in}(t) - u_{out}(t), \quad m(0) = m_0, \quad (2.7)$$

where $\mathbf{1}_p$ is a p -dimensional vector of ones and m_0 the initial mass. Note that the mass can also be computed from the numbers of moles as

$$m(t) = \mathbf{1}_S^T \mathbf{M}_w \mathbf{n}(t), \quad (2.8)$$

which indicates that Eqs. (2.6) and (2.7) are linearly dependent. Hence, the continuity equation is not needed per se, but it is often useful to express the mass as a function of the flows rather than the numbers of moles. The material balance equations for a homogeneous reaction system is illustrated in Example. 2.3.

Semi-batch reactor

If the reaction system is operated in a semi-batch reactor with p inlets, Eq. (2.6) reduces to

$$\dot{\mathbf{n}}(t) = \mathbf{N}^T \mathbf{r}_v(t) + \mathbf{W}_{in} \mathbf{u}_{in}(t), \quad \mathbf{n}(0) = \mathbf{n}_0. \quad (2.9)$$

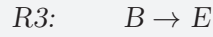
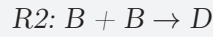
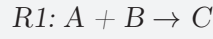
Batch reactor

Similarly for a batch reactor, Eq. (2.6) reduces to:

$$\dot{\mathbf{n}}(t) = \mathbf{N}^T \mathbf{r}_v(t), \quad \mathbf{n}(0) = \mathbf{n}_0. \quad (2.10)$$

Example 2.3 (Acetoacetylation of Pyrrole)

This reaction system consists of seven species involved in four independent reactions. The primary reaction (R1) between pyrrole (A) and diketene (B) produces the desired 2-acetoacetyl pyrrole (C). The side reactions includes the conversion (R2) of diketene to dehydroacetic acid (D), the oligomerization (R3) of diketene to oligomers (E) and the reaction (R4) of diketene and acetoacetyl pyrrole giving a by-product (F). Reactions R1, R2 and R4 are catalyzed by pyridine (K). The reaction network for this system can be represented as:



The stoichiometric matrix \mathbf{N} for this reaction system is given by:

$$\mathbf{N} = \begin{bmatrix} -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Consider the pyrrole system in a reactor with a single inlet and an outlet stream. The inlet stream feeds species A, B and K, and the species A, B and K are present initially. The material balance equations for all the seven species can be written as:

$$\begin{aligned} \dot{n}_A(t) &= -V(t)r_1(t) + w_{in,A}u_{in}(t) - \omega(t)n_A(t) \\ \dot{n}_B(t) &= -V(t)r_1(t) - 2V(t)r_2(t) - V(t)r_3(t) - V(t)r_4(t) + w_{in,B}u_{in}(t) - \omega(t)n_B(t) \\ \dot{n}_C(t) &= V(t)r_1(t) - V(t)r_4(t) - \omega(t)n_C(t) \\ \dot{n}_D(t) &= V(t)r_2(t) - \omega(t)n_D(t) \\ \dot{n}_E(t) &= V(t)r_3(t) - \omega(t)n_E(t) \\ \dot{n}_F(t) &= V(t)r_4(t) - \omega(t)n_F(t) \\ \dot{n}_K(t) &= w_{in,K}u_{in}(t) - \omega(t)n_K(t) \end{aligned}$$

The vector of initial conditions is given by, $\mathbf{n}_0 = [n_{A0}, n_{B0}, 0, 0, 0, 0, n_{K0}]^T$

The S -dimensional representation of the system given in Eq. (2.6) may contain redundant states. We define the dimensionality of the reaction system d as the number of rate processes affecting the reaction system, that is,

$$d = \# \text{ of endogenous inputs (reactions)} + \# \text{ of exogenous inputs (inlets and outlet)} \quad (2.11)$$

The dimensionality of the reaction system under different operating conditions is listed in Table 2.1

Table 2.1 Material balance equations for homogeneous reaction system

Reactor Type	Material Balance Equation	Dimensionality, d
Open with inlets and outlet	$\dot{\mathbf{n}}(t) = \mathbf{N}^T \mathbf{r}_v(t) + \mathbf{W}_{in} \mathbf{u}_{in}(t) - \omega(t) \mathbf{n}(t)$	$R + p + 1$
Semi-Batch	$\dot{\mathbf{n}}(t) = \mathbf{N}^T \mathbf{r}_v(t) + \mathbf{W}_{in} \mathbf{u}_{in}(t)$	$R + p$
Batch	$\dot{\mathbf{n}}(t) = \mathbf{N}^T \mathbf{r}_v(t)$	R

2.1.3 Invariant states

The concept of invariant relationships for homogeneous chemical reaction systems is presented next. We start by presenting the definition of an invariant relationship.

Definition 2.3 (Invariant relationship)

Any (linear or nonlinear) combination of the state variables $\mathbf{n}(t)$ is said to be an invariant relationship if its value remains constant over time.

This definition clearly shows that the invariant relationships constitute algebraic (i.e. static) relationships between the various states of the system. These invariant relationships have to be satisfied at all times, and therefore represent physical constraints regulating the evolution of the reaction system. The number of invariant relationships depends on the reactor type. Denoting their number by q , the invariant relationships can be expressed as:

$$\mathcal{I}(\mathbf{n}(t)) = \mathbf{0}_q, \quad (2.12)$$

with $\mathcal{I}(\cdot) = [\mathcal{I}_1(\cdot), \dots, \mathcal{I}_q(\cdot)]^T$ and $\mathcal{I}_i(\cdot)$ the i th invariant relationship of the system.

2.1.3.1 Invariant States – Reactors with Outlet

Procedure: To construct the invariants of a homogeneous reaction system with an outlet flow, construct the $S \times d$ matrix \mathbf{B} by concatenating the structural matrices of the system (\mathbf{N} and \mathbf{W}_{in}) along with the vector of initial conditions \mathbf{n}_0 , that is, $\mathbf{B} = [\mathbf{N}^T \ \mathbf{W}_{in} \ \mathbf{n}_0]$. The matrix \mathbf{B} is assumed to be of rank $d = R + p + 1$, which implies that $S \geq d$. Compute the $S \times q$ matrix \mathbf{P} such that the matrix lies in the null space of \mathbf{B}^T , that is,

$$\mathbf{B}^T \mathbf{P} = \mathbf{0}_{d \times q},$$

with $q := S - d = S - R - p - 1$ according to the rank nullity theorem [74]. Compute the q invariant relationships by pre-multiplying Eq. (2.6) with \mathbf{P}^T :

$$\mathbf{P}^T \mathbf{n}(t) = \mathbf{0}_q. \quad (2.13)$$

Remark 2.1

For a CSTR with constant volume and constant density, the dimensionality d is reduced to $R + p$ and the number of invariants is increased to $q = S - d = S - R - p$. The additional invariant for a CSTR can be derived from the conservation of mass, namely, $\mathbf{1}_S^T \mathbf{M}_w \mathbf{n}_0 - \mathbf{1}_S^T \mathbf{M}_w \mathbf{n}(t) = 0$.

Example 2.3 (Acetoacetylation of Pyrrole cont'd..)

Consider the pyrrole system in a CSTR with an inlet stream containing species A, B and K at the constant flow rate 2 g min^{-1} . The reactor also has an outlet stream with the constant flow rate 2 g min^{-1} . The volume and the density are constant and therefore the mass remains constant. The $S \times p$ inlet composition matrix \mathbf{W}_{in} is given by:

$$\mathbf{W}_{in} = \begin{bmatrix} 0.0060, 0.0064, 0, 0, 0, 0, 0.0008 \end{bmatrix}^T$$

The initial numbers of moles are $\mathbf{n}_0 = [2, 5, 0, 0, 0, 0, 0.5]$. The dimensionality of this system is $d = R + p = 4 + 1 = 5$ and the number of invariants, $q = S - d = 7 - 5 = 2$. The first invariant relationship is obtained according to the procedure described in Section 2.1.3.1:

$$n_K(t) - 0.0465 n_A(t) - 0.0814 n_B(t) - 0.1279 n_C(t) - 0.1628 n_D(t) - 0.0814 n_E(t) - 0.2093 n_F(t) = 0$$

In addition, as this reactor is a CSTR with a constant density, the second invariant relationship, obtained using the conservation of mass, is:

$$\mathbf{1}_S^T \mathbf{M}_w \mathbf{n}_0 - \mathbf{1}_S^T \mathbf{M}_w \mathbf{n}(t) = 0.$$

Remark 2.2

If $S \leq d$, the reaction system does not have any invariant relationships that can be constructed from the current form of \mathbf{B} . Section 2.1.3.3 gives a procedure for extracting invariants that uses flowrate information and a different definition of \mathbf{B} .

2.1.3.2 Invariant States – Reactors without Outlet

Procedure: The invariants of a homogeneous reaction system without outlet flow are obtained as follows: Construct the $S \times d$ matrix \mathbf{B} by concatenating the structural matrices of the system. For semi-batch conditions, $\mathbf{B} = [\mathbf{N}^T \mathbf{W}_{in}]$ with the rank $d = R + p$, whereas for batch conditions, $\mathbf{B} = \mathbf{N}^T$ with the rank of $d = R$. Compute the $S \times q$ matrix \mathbf{P} such that the matrix lies in the null space of \mathbf{B}^T ,

$$\mathbf{B}^T \mathbf{P} = \mathbf{0}_{d \times q},$$

that is, with $q = S - R - p$ for semi-batch conditions and $q = S - R$ for batch conditions. Compute the q invariant relationships as:

$$\mathbf{P}^T (\mathbf{n}(t) - \mathbf{n}_0) = \mathbf{0}_q. \quad (2.14)$$

Example 2.3 (Acetoacetylation of Pyrrole cont'd..)

Consider the acetoacetylation of pyrrole in a batch reactor with the initial conditions, $\mathbf{n}_0 = [2, 5, 0, 0, 0, 0, 0.5]$. The dimensionality for this operating mode is $d = R = 4$ and the number of invariant relationships $q = S - d = 3$. These relationships can be obtained using the procedure described in Section 2.1.3.2:

$$\begin{aligned} n_B(t) - 2n_A(t) - n_C(t) + 2n_D(t) + n_E(t) - (n_{B0} - 2n_{A0} - n_{C0} + 2n_{D0} + n_{E0}) &= 0 \\ n_A(t) + n_C(t) + n_F(t) - (n_{A0} + n_{C0} + n_{F0}) &= 0 \\ n_K(t) - n_{K0} &= 0 \end{aligned}$$

2.1.3.3 Invariant States using Flowrate Information

For semi-batch and open conditions, the $q = S - R - p$ and $q = S - R - p - 1$ invariants, respectively, can be increased to $q = S - R$ invariants if the flowrates $\mathbf{u}_{in}(t)$ and $u_{out}(t)$ are known. The procedure for computing this reduced number of invariants is as follows:

1. Write the numbers of moles in the *reaction-variant* form [63]:

$$\mathbf{n}^{RV}(t) := \mathbf{n}(t) - \mathbf{n}_0 - \int_0^t \mathbf{W}_{in} \mathbf{u}_{in}(t) dt + \int_0^t \omega(t) \mathbf{n}(t) dt. \quad (2.15)$$

2. Write the material balance equation in terms of the reaction-variant form:

$$\dot{\mathbf{n}}^{RV}(t) = \mathbf{N}^T \mathbf{r}_v(t). \quad (2.16)$$

3. Construct the $S \times R$ matrix $\mathbf{B} = \mathbf{N}^T$ with $d = R$.

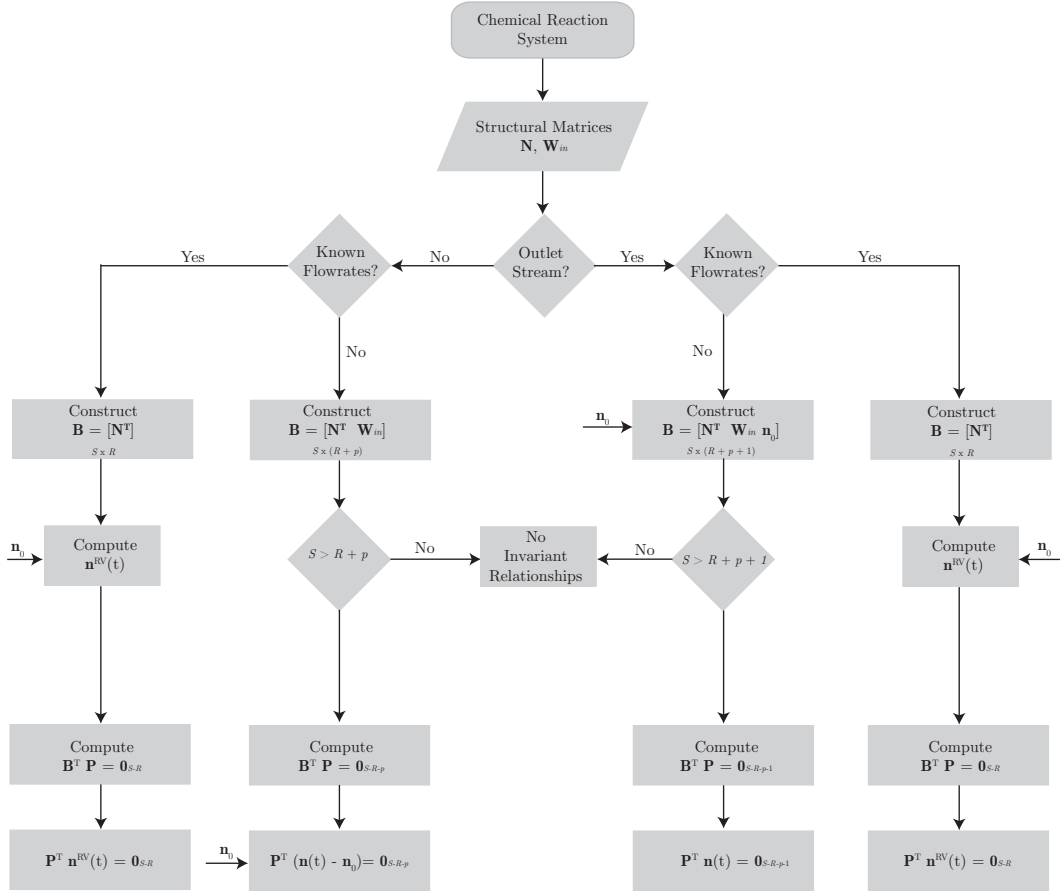


Figure 2.1 Schematic procedure for computing the invariant relationships of a chemical reaction system.

4. Compute the $S \times q$ matrix \mathbf{P} such that the matrix lies in the null space \mathbf{B}^T , that is,

$$\mathbf{B}^T \mathbf{P} = \mathbf{0}_{d \times q},$$

with $q := S - R$.

5. Compute the q invariants by pre-multiplying Eq. (2.16) with \mathbf{P}^T :

$$\mathbf{P}^T \mathbf{n}^{RV}(t) = \mathbf{0}_q. \quad (2.17)$$

Remark 2.3

The use of Eq. (2.15) involves integrating and interpolating noisy measured numbers of moles.

Fig. 2.1 summarizes the procedure for computing invariant relationships for reaction systems with and without an outlet stream.

2.1.4 Heat balance equations

Consider a homogeneous reaction system in an open non-isothermal reactor with a heating/cooling jacket. The heat balance around the reactor can be written as [2]:

$$\dot{Q}_r(t) = (-\Delta \mathbf{H}_r)^\top \mathbf{r}_v(t) + q_{ex}(t) + \check{\mathbf{T}}_{in}^\top \mathbf{u}_{in}(t) - \omega(t) Q_r(t), \quad Q_r(0) = Q_{r0}, \quad (2.18)$$

where $Q_r = m c_p T_r$ is the heat in the reactor, with T_r the reactor temperature and c_p the specific heat capacity of the reaction mixture, q_{ex} is the heat flow from the jacket to the reaction mixture, $\check{\mathbf{T}}_{in}$ the p -dimensional vector of specific heats of the inlet streams with $\check{T}_{in,j} = c_{p,in,j} T_{in,j}$ and $T_{in,j}$ the temperature of the j^{th} inlet, and $\Delta \mathbf{H}_r$ the R -dimensional vector of reaction enthalpies. Obviously, the heat-flow term $q_{ex}(t)$ depends on the reactor temperature, for example $q_{ex}(t) = UA (T_{ja}(t) - T_r(t))$ with the heat-transfer coefficient UA and the jacket temperature $T_{ja}(t)$, but this dependency is not included in Eq. (2.18). For simplicity, it is assumed that the inlet specific heat vector $\check{\mathbf{T}}_{in}$ is constant.

The heat balance for the jacket can be written as:

$$\dot{Q}_{ja}(t) = -q_{ex}(t) + \check{q}_{ja,in}(t), \quad Q_{ja}(0) = Q_{ja0}, \quad (2.19)$$

where $Q_{ja} = m_{ja} c_{p,ja} T_{ja}$ is the heat in the jacket, with $c_{p,ja}$ the specific heat capacity of the heating/cooling fluid, $\check{q}_{ja,in} = u_{ja}(t) c_{p,ja} (T_{ja,in} - T_{ja})$ is the net heat flow via the inlet flow of the heating/cooling fluid, u_{ja} the mass flow rate in the jacket and $T_{ja,in}$ the inlet temperature in the jacket.

Combined material and energy balances

The material balance equations (2.6) and the heat-balance equation inside the reactor (2.18) can be written in a compact form using the $(S + 1)$ -dimensional state vector:

$$\dot{\mathbf{z}}(t) = \mathcal{A} \mathbf{r}_v(t) + \mathbf{b} q_{ex}(t) + \mathcal{C} \mathbf{u}_{in}(t) - \omega(t) \mathbf{z}(t), \quad \mathbf{z}(0) = \mathbf{z}_0, \quad (2.20)$$

where $\mathbf{z}(t) = \begin{bmatrix} \mathbf{n}(t) \\ Q_r(t) \end{bmatrix}$, $\mathcal{A} = \begin{bmatrix} \mathbf{N}^\top \\ (-\Delta \mathbf{H}_r)^\top \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} \mathbf{0}_S \\ 1 \end{bmatrix}$, $\mathcal{C} = \begin{bmatrix} \mathbf{W}_{in} \\ \check{\mathbf{T}}_{in}^\top \end{bmatrix}$ and $\mathbf{z}_0 = \begin{bmatrix} \mathbf{n}_0 \\ Q_{r0} \end{bmatrix}$.

System invariants for combined material and energy balance equations

The procedure for deriving the system invariants can be extended to include the heat-balance equations along with the material balance equations. The procedure is the

following: Construct a matrix \mathbf{B} by concatenating the terms \mathcal{A} , \mathbf{b} , \mathcal{C} and \mathbf{z}_0 as $\mathbf{B} = [\mathcal{A} \ \mathbf{b} \ \mathcal{C} \ \mathbf{z}_0]$ of dimension $(S+1) \times (R+p+2)$. Compute an $(S+1) \times q$ matrix \mathbf{P} such that this matrix lies in the null space of the matrix \mathbf{B}^\top ,

$$\mathbf{B}^\top \mathbf{P} = \mathbf{0}_{d \times q}$$

where $q = (S+1) - (R+p+2)$. Pre-multiplying Eq. (2.20) by the matrix \mathbf{P}^\top results in the q invariant relationships for the combined material and energy balances of a homogeneous reaction system in an open reactor:

$$\mathbf{P}^\top \mathbf{z}(t) = \mathbf{0}_q \quad (2.21)$$

2.2 Heterogeneous Reaction Systems

Consider a two-phase chemical reaction system with phases G and L. The system has S species living in the S -dimensional set \mathcal{S} . Among these S species, S_f species are present in phase F and constitute the set \mathcal{S}_f , with $F \in \{G, L\}$, $f \in \{g, l\}$ and $S = S_g + S_l$. These S_f species are involved in R_f independent reactions in phase F. There are p_f independent inlet and one outlet stream in each phase.

2.2.1 Material balance equations

Under the assumption that the reactions are taking place only in the bulk of each phase and that the two phases are connected by p_m steady-state mass transfers, the differential mole balance equations for phase F, can be written as:

$$\dot{\mathbf{n}}_f(t) = \mathbf{N}_f^\top \mathbf{r}_{v,f}(t) + \mathbf{W}_{m,f} \boldsymbol{\zeta}_f(t) + \mathbf{W}_{in,f} \mathbf{u}_{in,f}(t) - \omega_f(t) \mathbf{n}_f(t), \quad \mathbf{n}_f(0) = \mathbf{n}_{f0}, \quad (2.22)$$

where \mathbf{n}_f is the S_f -dimensional vector of numbers of moles, \mathbf{N}_f is the $R_f \times S_f$ stoichiometric matrix, $\mathbf{r}_{v,f}(t) := V_f(t) \mathbf{r}_f(t)$ is the R_f -dimensional vector of reaction rates, $\mathbf{W}_{in,f}$ is the $S_f \times p_f$ inlet matrix expressing the composition of the inlets to phase F, $\mathbf{u}_{in,f}$ is the p_f -dimensional vector of inlet mass flow rates to phase F, $\omega_f(t) := \frac{u_{out,f}(t)}{m_f(t)}$ is the inverse residence time, with $u_{out,f}$ the outlet mass flowrate and m_f the mass of phase F, and \mathbf{n}_{f0} is the vector of initial numbers of moles. The p_m mass transfers are treated as pseudo inlets with unknown flowrates, where $\mathbf{W}_{m,f} = \mathbf{M}_{w,f}^{-1} \check{\mathbf{E}}_{m,f}$ is the $S_f \times p_m$ mass transfer matrix for phase F, $\check{\mathbf{E}}_{m,f} = [\check{\mathbf{e}}_{m,f}^1 \cdots \check{\mathbf{e}}_{m,f}^{p_m}]$ with $\check{\mathbf{e}}_{m,f}^j$ being the S_f -dimensional vector with the elements corresponding to the j th transferring species equal to unity and the other elements equal to zero and $\boldsymbol{\zeta}_f$ is the p_m -dimensional vector of mass-transfer rates expressed in $\frac{\text{mass}}{\text{time}}$ units. The mass-transfer rate $\zeta_{f,j}$ of the j th species is positive if the j th species is added into phase F due to mass transfer and it is negative if that species leaves the phase due to mass transfer. This implies that $\boldsymbol{\zeta}_g = -\boldsymbol{\zeta}_l$. The mass in phase F can be calculated from the numbers of moles in that phase as

$\mathbf{m}_f(t) = \mathbf{1}_s^T \mathbf{M}_{w,f} \mathbf{n}_f(t)$. Alternatively, the mass can be computed using the continuity equation:

$$\dot{m}_f(t) = \mathbf{1}_{p_m}^T \boldsymbol{\zeta}_f(t) + \mathbf{1}_{p_f}^T \mathbf{u}_{in,f}(t) - u_{out,f}(t), \quad m_f(0) = m_{f0}. \quad (2.23)$$

Semi-batch reactor

If the heterogeneous reaction system is operated in a semi-batch reactor, the material balance equation reduces to:

$$\dot{\mathbf{n}}_f(t) = \mathbf{N}_f^T \mathbf{r}_{v,f}(t) + \mathbf{W}_{m,f} \boldsymbol{\zeta}_f(t) + \mathbf{W}_{in,f} \mathbf{u}_{in,f}(t), \quad \mathbf{n}_f(0) = \mathbf{n}_{f0}. \quad (2.24)$$

Batch reactor

The material balance equation for the reaction system operated in a batch reactor can be written as:

$$\dot{\mathbf{n}}_f(t) = \mathbf{N}_f^T \mathbf{r}_{v,f}(t) + \mathbf{W}_{m,f} \boldsymbol{\zeta}_f(t), \quad \mathbf{n}_f(0) = \mathbf{n}_{f0}. \quad (2.25)$$

2.2.2 Invariant states

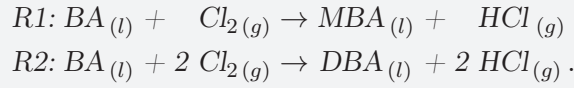
Similar to the procedure for homogeneous reaction systems, the invariant relationships can be derived for each phase F of a heterogeneous reaction system. The dimensionality, d_f and the $q_f = S_f - d_f$ invariant relationships for phase F of a heterogeneous reaction system operated in different operating modes are shown in Table 2.2.

Table 2.2 Algebraic constraints under different operating modes for phase F of a heterogeneous reaction system.

Reactor	d_f	Constraint derivation	Invariants
Open with inlets and outlet	$R_f + p_m + p_f + 1$	$\mathbf{P}_f^T [\mathbf{N}_f^T \mathbf{W}_{m,f} \mathbf{W}_{in,f} \mathbf{n}_{f0}] = \mathbf{0}_{q_f \times d_f}$	$\mathbf{P}_f^T \mathbf{n}_f(t) = \mathbf{0}_{q_f}$
Semi-batch	$R_f + p_m + p_f$	$\mathbf{P}_f^T [\mathbf{N}_f^T \mathbf{W}_{m,f} \mathbf{W}_{in,f}] = \mathbf{0}_{q_f \times d_f}$	$\mathbf{P}_f^T (\mathbf{n}_f(t) - \mathbf{n}_{f0}) = \mathbf{0}_{q_f}$
Batch	$R_f + p_m$	$\mathbf{P}_f^T [\mathbf{N}_f^T \mathbf{W}_{m,f}] = \mathbf{0}_{q_f \times d_f}$	$\mathbf{P}_f^T (\mathbf{n}_f(t) - \mathbf{n}_{f0}) = \mathbf{0}_{q_f}$

Example 2.4 (Chlorination of Butanoic Acid)

The chlorination of butanoic acid (BA) is a gas-liquid reaction system involving two reactions in the liquid phase that consume chlorine (Cl_2) dissolved from the gas phase. The first reaction produces α -mono-chloro-butanoic acid (MBA) and hydrochloric acid (HCl), while the second reaction produces the side product α -di-chloro-butanoic acid (DBA) and HCl. Due to its high volatility, HCl can be found in both phases. Hence, this reaction system has two species in the gas phase $\{\text{Cl}_2, \text{HCl}\}$ and five species in the liquid phase $\{\text{Cl}_2, \text{BA}, \text{MBA}, \text{HCl}, \text{DBA}\}$. Ethanol is generally used as solvent for the liquid phase. The reaction can be represented by the following scheme:



Assume that the reaction system is operated in a semi-batch reactor, with Cl_2 fed in the gas phase and BA present initially in the liquid phase. The material balance equation for phase G can be written as:

$$\begin{aligned} \dot{n}_{\text{Cl}_2,g}(t) &= \zeta_{\text{Cl}_2,g}(t) + w_{in,\text{Cl}_2,g} u_{in,g}(t) & n_{\text{Cl}_2,g}(0) &= 0 \\ \dot{n}_{\text{HCl},g}(t) &= \zeta_{\text{HCl},g}(t) & n_{\text{HCl},g}(0) &= 0 \end{aligned}$$

For gas phase G, the mass transfer rate $\zeta_{\text{Cl}_2,g}(t)$ is negative, while $\zeta_{\text{HCl},g}(t)$ is positive. The material balance equations for phase L can be written as:

$$\begin{aligned} \dot{n}_{\text{BA},l}(t) &= -V_l(t)r_{l,1}(t) - V_l(t)r_{l,2}(t) & n_{\text{BA},l}(0) &= n_{\text{BA}0} \\ \dot{n}_{\text{Cl}_2,l}(t) &= -V_l(t)r_{l,1}(t) - 2V_l(t)r_{l,2}(t) + \zeta_{\text{Cl}_2,l}(t) & n_{\text{Cl}_2,l}(0) &= 0 \\ \dot{n}_{\text{MBA},l}(t) &= V_l(t)r_{l,1}(t) & n_{\text{MBA},l}(0) &= 0 \\ \dot{n}_{\text{HCl},l}(t) &= V_l(t)r_{l,1}(t) + 2V_l(t)r_{l,2}(t) + \zeta_{\text{HCl},l}(t) & n_{\text{HCl},l}(0) &= 0 \\ \dot{n}_{\text{DBA},l}(t) &= V_l(t)r_{l,2}(t) & n_{\text{DBA},l}(0) &= 0 \end{aligned}$$

For phase L, the mass transfer rate $\zeta_{\text{Cl}_2,l}(t)$ is positive, while $\zeta_{\text{HCl},l}(t)$ is negative.

2.3 Summary

This chapter has derived a relationship between the number of species and the maximum number of independent reactions for a chemical reaction system. The material and energy balance equations for a homogeneous chemical reaction system in a generic open reactor with inlet and outlet streams have also been explained. This chapter has introduced the concept of invariant relationships, and a procedure for deriving these invariant relationships for a homogeneous reaction system has been developed. The material balance equations for a multi-phase heterogeneous reaction system have been explained. Finally, the invariant relationships in each phase of the heterogeneous reaction systems have been derived.

Chapter 3

Vessel Extents Representation

The material balance equations in terms of the numbers of moles introduced in the previous chapter are functions of all the rate processes (reactions, mass transfers, inlets and outlets). The presence of all these phenomena, and in particular their interactions, complicates the analysis of chemical reactions. The analysis would become more efficient if the effects of these various phenomena could somehow be separated and each phenomenon is investigated individually.

Sections 3.1.1 and 3.1.2 briefly summarize the two-way and three-way decoupled representations of chemical reaction system in terms of variant and invariant states. Section 3.1.3 introduces the representation in terms of *vessel extents* for a homogeneous reaction system. A linear transformation that brings the representation of a homogeneous reaction system from numbers of moles to vessel extents is also introduced. The conditions under which the transformation between these representations are possible is explained in detail. The concept of vessel extents is then extended to reaction systems where the heat transfer between the reactor and its jacket is also modeled. This requires the introduction of an additional vessel extent for heat exchange. Section 3.2.1 extends the representation in terms of vessel extents for a heterogeneous reaction system.

3.1 Homogeneous Reaction Systems

In this section, the formal definitions of reaction variants and reaction invariants are introduced. The various decoupled representations for chemical reaction systems previously introduced in the literature are briefly discussed in Section 3.1.1 and Section 3.1.2. In Section 3.1.3, we introduce a representation of a chemical reaction system in terms of vessel extents.

Definition 3.1 (Reaction invariants)

State variables that are not affected by reactions are termed reaction invariants.

Definition 3.2 (Reaction variants)

All other state variables that are affected by chemical reactions are called reaction variants.

3.1.1 Two-way representation of chemical reaction systems

For a homogeneous reaction system, Asbjørnsen et al. [7, 8, 33] introduced a representation in terms of the reaction-variant states $\mathbf{y}_r(t)$ and reaction-invariant states $\mathbf{y}_{iv}(t)$. The resulting dynamical system contains the R state variables \mathbf{y}_r that depend on the reactions and the $q = S - R$ state variables \mathbf{y}_{iv} that do not:

$$\begin{aligned}\dot{\mathbf{y}}_r(t) &= \mathbf{r}_v(t) + \mathbf{N}^{\text{T}+} \mathbf{W}_{in} \mathbf{u}_{in}(t) - \omega(t) \mathbf{y}_r(t), & \mathbf{y}_r(0) &= \mathbf{N}^{\text{T}+} \mathbf{n}_0, \\ \dot{\mathbf{y}}_{iv}(t) &= \mathbf{P}^+ \mathbf{W}_{in} \mathbf{u}_{in}(t) - \omega(t) \mathbf{y}_{iv}(t), & \mathbf{y}_{iv}(0) &= \mathbf{P}^+ \mathbf{n}_0,\end{aligned}\quad (3.1)$$

where the matrix \mathbf{P} of dimension $S \times q$ describes the null space of the stoichiometric matrix, that is, $\mathbf{N}\mathbf{P} = \mathbf{0}_{R \times q}$. More specifically, one sees that the reaction variants are decoupled with respect to the reaction rates, that is, $y_{r,i}(t)$ depends on $r_{v,i}(t)$, but not on the other reaction rates:

$$\dot{y}_{r,i}(t) = r_{v,i}(t) + (\mathbf{N}^{\text{T}+} \mathbf{W}_{in})_i \mathbf{u}_{in}(t) - \omega(t) y_{r,i}(t), \quad y_{r,i}(0) = (\mathbf{N}^{\text{T}+})_i \mathbf{n}_0, \quad (3.2)$$

where $(\cdot)_i$ represents the i th row of the matrix (\cdot) . Let $\mathbf{y}(t)$ represent the concatenated matrix $\begin{bmatrix} \mathbf{y}_r(t) \\ \mathbf{y}_{iv}(t) \end{bmatrix}$.

Transformation from \mathbf{y} to \mathbf{n} : The transformation from $\mathbf{y}_r(t)$ and $\mathbf{y}_{iv}(t)$ to the number of moles can be performed using the relationship:

$$\mathbf{n}(t) = \mathbf{N}^{\text{T}} \mathbf{y}_r(t) + \mathbf{P}^{\text{T}} \mathbf{y}_{iv}(t). \quad (3.3)$$

Transformation from \mathbf{n} to \mathbf{y} : If $\text{rank}(\mathbf{N}) = R$, Asbjørnsen et al. [7, 8, 33] introduced a linear transformation for converting the numbers of moles to the reaction variant and reaction invariant states. This transformation \mathcal{T}_{asb} involves the stoichiometric matrix \mathbf{N} and its null space \mathbf{P} , that is :

$$\begin{bmatrix} \mathbf{y}_r(t) \\ \mathbf{y}_{iv}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{R} \\ \mathbf{Q} \end{bmatrix} \mathbf{n}(t) := \mathcal{T}_{asb} \mathbf{n}(t), \quad \mathcal{T}_{asb} = [\mathbf{N}^{\text{T}} \ \mathbf{P}]^{-1}. \quad (3.4)$$

It follows from $\mathbf{N}\mathbf{P} = \mathbf{0}_{R \times q}$ and $\mathcal{T}_{asb} \mathcal{T}_{asb}^{-1} = \begin{bmatrix} \mathbf{R} \\ \mathbf{Q} \end{bmatrix} [\mathbf{N}^{\text{T}} \ \mathbf{P}] = \begin{bmatrix} \mathbf{I}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix}$ that $\mathbf{R} = \mathbf{N}^{\text{T}+}$ and $\mathbf{Q} = \mathbf{P}^+$, where $\mathbf{N}^{\text{T}+}$ and \mathbf{P}^+ represent the Moore-Penrose pseudo-inverse of \mathbf{N}^{T} and \mathbf{P} , respectively.

Remark 3.1

Since $S > R$ from Eq. (2.5), the rank of the $R \times S$ matrix \mathbf{N} is R . This implies that the transformation \mathcal{T}_{asb} from $\mathbf{n}(t)$ to $\mathbf{y}(t)$ is always possible.

Remark 3.2

The quantities \mathbf{y}_r are reaction and flow variants and not solely reaction variants. \mathbf{y}_{iv} are reaction invariants but flow variants, hence not truly invariants. Note that \mathbf{y}_r are

pure reaction variants and \mathbf{y}_{iv} are true system invariants *only for batch reactors*, i.e., with $\mathbf{u}_{in}(t) = \mathbf{0}_p$ and $\omega(t) = 0$, for which one can write:

$$\begin{aligned}\dot{\mathbf{y}}_r(t) &= \mathbf{r}_v(t), & \mathbf{y}_r(0) &= \mathbf{N}^{\text{T}^+} \mathbf{n}_0, \\ \mathbf{y}_{iv}(t) &= \mathbf{P}^+ \mathbf{n}_0.\end{aligned}\tag{3.5}$$

Remark 3.3

The quantities \mathbf{y}_r and \mathbf{y}_{iv} are therefore abstract mathematical quantities with no physical meaning.

3.1.2 Three-way representation of chemical reaction systems

Amrhein et al. [70] introduced a representation of homogeneous reaction systems in terms of three quantities (i) reaction variants, (ii) flow variants, and (iii) reaction and inlet-flow invariants. In this representation, the reaction system contains R state variables that depend on reactions, p state variables that depend on the inlet flows and $S - R - p$ states that do not depend on reactions and inlet flows:

$$\begin{aligned}\dot{\mathbf{w}}_r(t) &= \mathbf{r}_v(t) - \omega(t) \mathbf{w}_r(t), & \mathbf{w}_r(0) &= \mathbf{S}^{\text{T}} \mathbf{n}_0, \\ \dot{\mathbf{w}}_{in}(t) &= \mathbf{u}_{in}(t) - \omega(t) \mathbf{w}_{in}(t), & \mathbf{w}_{in}(0) &= \mathbf{M}^{\text{T}} \mathbf{n}_0, \\ \dot{\mathbf{w}}_{iv}(t) &= -\omega(t) \mathbf{w}_{iv}(t) & \mathbf{w}_{iv}(0) &= \mathbf{Q}^{\text{T}} \mathbf{n}_0,\end{aligned}\tag{3.6}$$

where the matrices \mathbf{S} , \mathbf{M} and \mathbf{Q} are of dimensions $S \times R$, $S \times p$ and $S \times (S - R - p)$ respectively. Let $\mathbf{w}(t)$ represent the concatenated matrix $\begin{bmatrix} \mathbf{w}_r(t) \\ \mathbf{w}_{in}(t) \\ \mathbf{w}_{iv}(t) \end{bmatrix}$.

Transformation from \mathbf{w} to \mathbf{n} : The number of moles $\mathbf{n}(t)$ can be reconstructed from $\mathbf{w}_r(t)$, $\mathbf{w}_{in}(t)$ and $\mathbf{w}_{iv}(t)$ using:

$$\mathbf{n}(t) = \mathbf{N}^{\text{T}} \mathbf{w}_r(t) + \mathbf{W}_{in} \mathbf{w}_{in}(t) + \mathbf{Q} \mathbf{w}_{iv}(t).\tag{3.7}$$

Transformation from \mathbf{n} to \mathbf{w} : If $\text{rank}[\mathbf{N}^{\text{T}} \ \mathbf{W}_{in}] = R + p$, the linear transformation \mathcal{T}_{amr} transforms $\mathbf{n}(t)$ to $\mathbf{w}(t)$ and can be represented as:

$$\begin{bmatrix} \mathbf{w}_r(t) \\ \mathbf{w}_{in}(t) \\ \mathbf{w}_{iv}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{S}^{\text{T}} \\ \mathbf{M}^{\text{T}} \\ \mathbf{Q}^{\text{T}} \end{bmatrix} \mathbf{n}(t) := \mathcal{T}_{amr} \mathbf{n}(t), \quad \mathcal{T}_{amr} = [\mathbf{N}^{\text{T}} \ \mathbf{W}_{in} \ \mathbf{P}]^{-1},\tag{3.8}$$

where the matrix \mathbf{P} of dimension $q = S - R - p$ describes the null space of the matrix $[\mathbf{N}^{\text{T}} \ \mathbf{W}_{in}]$, that is, $\mathbf{P}^{\text{T}} [\mathbf{N}^{\text{T}} \ \mathbf{W}_{in}] = \mathbf{0}_{q \times R+p}$.

Remark 3.4

Since the matrix $[\mathbf{N}^T \ \mathbf{W}_{in}]$ has $R + p$ independent columns, its rank is equal to $\min(S, R + p)$. If $S < R + p$, the transformation \mathcal{T}_{amr} cannot be applied, i.e., it is not possible to transform $\mathbf{n}(t)$ to $\mathbf{w}(t)$. The transformation from $\mathbf{w}(t)$ to $\mathbf{n}(t)$ using Eq. (3.7) is however always possible.

Remark 3.5

The transformed variable introduced in this section do not have a physical meaning. Furthermore, the transformed variables become pure reaction variants even for a semi-batch reactor when $\omega(t) = 0$,

$$\begin{aligned} \dot{\mathbf{w}}_r(t) &= \mathbf{r}_v(t), & \mathbf{w}_r(0) &= \mathbf{S}^T \mathbf{n}_0, \\ \dot{\mathbf{w}}_{in}(t) &= \mathbf{u}_{in}(t), & \mathbf{w}_{in}(0) &= \mathbf{M}^T \mathbf{n}_0, \\ \dot{\mathbf{w}}_{iv}(t) &= \mathbf{Q}^T \mathbf{n}_0. \end{aligned} \quad (3.9)$$

3.1.3 Vessel extents representation

In this section, we introduce an alternative representation of homogeneous reaction systems in terms of vessel extents. The following definitions of vessel extents are adapted from the work of Bhatt [13]:

Definition 3.3 (Vessel Extent of i th Reaction, $x_{r,i}$)

The vessel extent of the i th reaction $x_{r,i}(t)$ is defined as the numbers of moles that is produced by the i th reaction that are still inside the reactor at time t . Hence, one can write:

$$\dot{x}_{r,i}(t) = r_{v,i}(t) - \omega(t)x_{r,i}(t), \quad x_{r,i}(0) = 0.$$

The term $-\omega(t)x_{r,i}(t)$ accounts for the material produced by the i th reaction that is removed from the reactor by the outlet stream. Note that, in the absence of an outlet stream, the definition of a vessel extent reduces to the batch extent of reaction $\xi_{r,i}$.

$$\dot{\xi}_{r,i}(t) = r_{v,i}(t), \quad \xi_{r,i}(0) = 0.$$

Definition 3.4 (Vessel Extent of j th Inlet, $x_{in,j}$)

The vessel extent of j th inlet $x_{in,j}(t)$ is defined as the mass added by the j th inlet stream that is still inside the reactor at time t . Hence,

$$\dot{x}_{in,j}(t) = u_{in,j}(t) - \omega(t)x_{in,j}(t), \quad x_{in,j}(0) = 0.$$

The term $-\omega(t)x_{in,j}(t)$ accounts for the material added by the j th inlet that has left the reactor by the outlet stream.

Definition 3.5 (Vessel Extent of Initial Condition, x_{ic})

The vessel extent of initial conditions $x_{ic}(t)$ is the fraction of the initial conditions that is still inside the reactor at time t . Hence,

$$\dot{x}_{ic}(t) = -\omega(t) x_{ic}(t), \quad x_{ic}(0) = 1.$$

The chemical reaction system can then be represented in terms of vessel extents as:

$$\dot{x}_{r,i}(t) = r_{v,i}(t) - \omega(t) x_{r,i}(t), \quad x_{r,i}(0) = 0, \quad i = 1, \dots, R, \quad (3.10a)$$

$$\dot{x}_{in,j}(t) = u_{in,j}(t) - \omega(t) x_{in,j}(t), \quad x_{in,j}(0) = 0, \quad j = 1, \dots, p, \quad (3.10b)$$

$$\dot{x}_{ic}(t) = -\omega(t) x_{ic}(t), \quad x_{ic}(0) = 1, \quad (3.10c)$$

$$\mathbf{x}_{iv}(t) = \mathbf{0}_q, \quad (3.10d)$$

where $x_{r,i}(t)$ is the extent of the i th reaction at time t expressed in mol, $x_{in,j}(t)$ the extent of the j th inlet flow at time t expressed in g, $x_{ic}(t)$ is the extent of initial conditions and $\mathbf{x}_{iv}(t)$ the vector of invariants at time t . Let $\mathbf{x}(t)$ represent the concatenated vector

$$\begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{in}(t) \\ x_{ic}(t) \\ \mathbf{x}_{iv}(t) \end{bmatrix}.$$

Transformation from \mathbf{x} to \mathbf{n} : The numbers of moles $\mathbf{n}(t)$ can be obtained from $\mathbf{x}(t)$ as:

$$\mathbf{n}(t) = \mathbf{N}^T \mathbf{x}_r(t) + \mathbf{W}_{in} \mathbf{x}_{in}(t) + \mathbf{n}_0 x_{ic}(t). \quad (3.11)$$

Transformation from \mathbf{n} to \mathbf{x} : If $\text{rank}([\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0]) = d$, then the linear transformation \mathcal{T} transforms $\mathbf{n}(t)$ to $\mathbf{x}(t)$. The transformation can be represented as:

$$\begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{in}(t) \\ x_{ic}(t) \\ \mathbf{x}_{iv}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{R} \\ \mathbf{F} \\ \mathbf{i}^T \\ \mathbf{Q} \end{bmatrix} \mathbf{n}(t) := \mathcal{T} \mathbf{n}(t), \quad (3.12)$$

and brings the dynamic model (2.6) to the following decoupled form:

$$\begin{aligned} \dot{\mathbf{x}}_r(t) &= \underbrace{\mathbf{R}\mathbf{N}^T}_{\mathbf{I}_R} \mathbf{r}_v(t) + \underbrace{\mathbf{R}\mathbf{W}_{in}}_{\mathbf{0}_{R \times p}} \mathbf{u}_{in}(t) - \omega(t) \mathbf{x}_r(t), & \mathbf{x}_r(0) &= \underbrace{\mathbf{R} \mathbf{n}_0}_{\mathbf{0}_R}, \\ \dot{\mathbf{x}}_{in}(t) &= \underbrace{\mathbf{F}\mathbf{N}^T}_{\mathbf{0}_{p \times R}} \mathbf{r}_v(t) + \underbrace{\mathbf{F}\mathbf{W}_{in}}_{\mathbf{I}_p} \mathbf{u}_{in}(t) - \omega(t) \mathbf{x}_{in}(t), & \mathbf{x}_{in}(0) &= \underbrace{\mathbf{F} \mathbf{n}_0}_{\mathbf{0}_p}, \\ \dot{x}_{ic}(t) &= \underbrace{\mathbf{i}^T \mathbf{N}^T}_{\mathbf{0}_R^T} \mathbf{r}_v(t) + \underbrace{\mathbf{i}^T \mathbf{W}_{in}}_{\mathbf{0}_p^T} \mathbf{u}_{in}(t) - \omega(t) x_{ic}(t), & x_{ic}(0) &= \underbrace{\mathbf{i}^T \mathbf{n}_0}_1, \\ \dot{\mathbf{x}}_{iv}(t) &= \underbrace{\mathbf{Q}\mathbf{N}^T}_{\mathbf{0}_{q \times R}} \mathbf{r}_v(t) + \underbrace{\mathbf{Q}\mathbf{W}_{in}}_{\mathbf{0}_{q \times p}} \mathbf{u}_{in}(t) - \omega(t) \mathbf{x}_{iv}(t), & \mathbf{x}_{iv}(0) &= \underbrace{\mathbf{Q} \mathbf{n}_0}_{\mathbf{0}_q}, \end{aligned} \quad (3.13)$$

where \mathbf{R} , \mathbf{F} and \mathbf{Q} are matrices of dimension $R \times S$, $p \times S$, and $q \times S$, respectively, and \mathbf{i} is a S -dimensional vector, with $q = S - d$ being the number of invariant quantities.

The linear transformation \mathcal{T} is:

$$\mathcal{T} = [\mathbf{N}^T \ \mathbf{W}_{in} \ \mathbf{n}_0 \ \mathbf{P}]^{-1}, \quad (3.14)$$

where the matrix \mathbf{P} describes the q -dimensional null space of the matrix $[\mathbf{N}^T \ \mathbf{W}_{in} \ \mathbf{n}_0]^T$. The transformation \mathcal{T} gives the conditions shown under the braces in equation (3.13), namely:

$$\begin{bmatrix} \mathbf{R} \\ \mathbf{F} \\ \mathbf{i}^T \\ \mathbf{Q} \end{bmatrix} [\mathbf{N}^T \ \mathbf{W}_{in} \ \mathbf{n}_0 \ \mathbf{P}] = \begin{bmatrix} \mathbf{I}_R & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix}. \quad (3.15)$$

It follows from \mathbf{P} being orthogonal to $[\mathbf{N}^T \ \mathbf{W}_{in} \ \mathbf{n}_0]$ and $\mathbf{Q}\mathbf{P} = \mathbf{I}_q$ that $\mathbf{Q} = \mathbf{P}^+$. Furthermore, $\mathbf{N}^T \mathbf{R} + \mathbf{W}_{in} \mathbf{F} + \mathbf{n}_0 \mathbf{i}^T + \mathbf{P}\mathbf{P}^+ = \mathbf{I}_S$, where $\mathbf{N}^T \mathbf{R}$ represents the R -dimensional reaction subspace, $\mathbf{W}_{in} \mathbf{F}$ the p -dimensional inlet subspace, $\mathbf{n}_0 \mathbf{i}^T$ the one-dimensional subspace describing the contribution of the initial conditions, and $\mathbf{P}\mathbf{P}^+$ the q -dimensional invariant subspace as shown in Fig. 3.1. All subspaces add up to the S -dimensional species space \mathbb{R}^S . Note that the invariant subspace is *orthogonal* to the other subspaces by construction, while the other subspaces are typically not orthogonal to each other. An alternative transformation from $\mathbf{n}(t)$ to $\mathbf{x}(t)$ has been introduced by Bhatt [13].

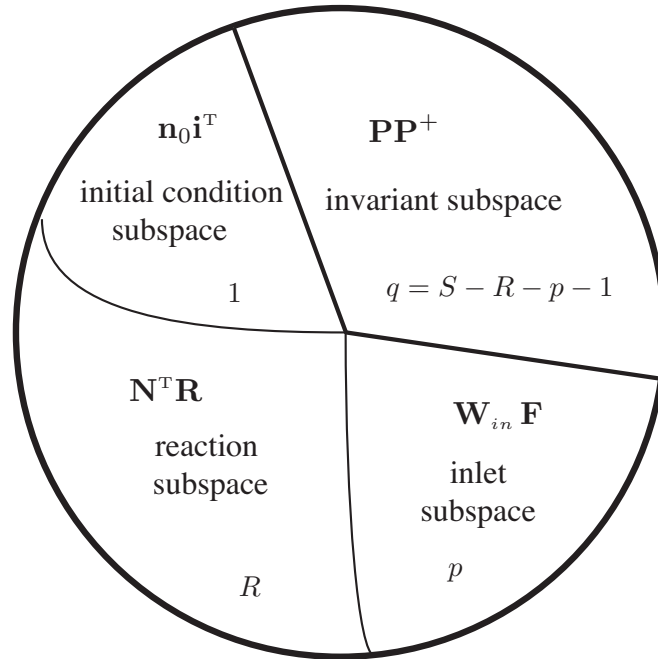


Figure 3.1 Decomposition of the S -dimensional space of numbers of moles into an R -dimensional reaction subspace, a p -dimensional inlet subspace, a one-dimensional subspace describing the contribution of the initial conditions, and a q -dimensional invariant subspace.

Remark 3.6

If the matrix $[\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0]$ has d independent columns, its rank is equal to $\min(S, d)$. As introduced in the previous chapter, $S > R$. Hence, to apply the transformation \mathcal{T} , one must necessarily have $S \geq d$.

Example 3.1 (Acetoacetylation of Pyrrole)

Consider the acetoacetylation of pyrrole system with four independent reactions given in Example 2.3. The reaction system is operated in a reactor with an inlet stream containing species A, B and K at the constant flow rate 2 g min^{-1} . The reactor also has an outlet stream with a constant flow rate 1.5 g min^{-1} . The $S \times p$ inlet composition matrix \mathbf{W}_{in} is given by:

$$\mathbf{W}_{in} = \begin{bmatrix} 0.0060, 0.0064, 0, 0, 0, 0, 0.0008 \end{bmatrix}^T \quad (3.16)$$

The initial numbers of moles are $\mathbf{n}_0 = [2, 5, 0, 0, 0, 0, 0.5]$. The system representation in terms of vessel extents for this system can be written as:

$$\begin{aligned} \dot{x}_{r,1}(t) &= V(t)r_1(t) - \omega(t)x_{r,1}(t), & x_{r,1}(0) &= 0, \\ \dot{x}_{r,2}(t) &= V(t)r_2(t) - \omega(t)x_{r,2}(t), & x_{r,2}(0) &= 0, \\ \dot{x}_{r,3}(t) &= V(t)r_3(t) - \omega(t)x_{r,3}(t), & x_{r,3}(0) &= 0, \\ \dot{x}_{r,4}(t) &= V(t)r_4(t) - \omega(t)x_{r,4}(t), & x_{r,4}(0) &= 0, \\ \dot{x}_{in}(t) &= u_{in}(t) - \omega(t)x_{in}(t), & x_{in}(0) &= 0, \\ \dot{x}_{ic}(t) &= -\omega(t)x_{ic}(t), & x_{ic}(0) &= 1, \\ x_{iv}(t) &= 0. \end{aligned}$$

The numbers of moles can be reconstructed using Eq. (3.11). The matrix $[\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0]$ of dimension (7×6) has rank $d = R + p + 1 = 6$. Therefore, the transformation matrix \mathcal{T} can be used to convert the numbers of moles to vessel extents.

Next, the representations in terms of extents for three special cases, namely, batch reactors, semi-batch reactors, and CSTRs are discussed.

3.1.3.1 Batch reactors

The vessel-extent representation for a batch reactor with $p = 0$ and $\omega(t) = 0$ can be written as:

$$\dot{x}_{r,i}(t) = r_{v,i}(t), \quad x_{r,i}(0) = 0, \quad i = 1, \dots, R, \quad (3.17a)$$

$$\mathbf{x}_{iv}(t) = \mathbf{0}_q, \quad (3.17b)$$

with $q = S - R$.

Transformation from \mathbf{x} to \mathbf{n} : The numbers of moles $\mathbf{n}(t)$ can be expressed as:

$$\mathbf{n}(t) = \mathbf{N}^T \mathbf{x}_r(t) + \mathbf{n}_0. \quad (3.18)$$

Transformation from \mathbf{n} to \mathbf{x} : For a batch reactor,

$$\begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{iv}(t) \end{bmatrix} := \mathcal{T} (\mathbf{n}(t) - \mathbf{n}_0), \quad (3.19)$$

where, $\mathcal{T} = [\mathbf{N}^T \ \mathbf{P}]^{-1}$ and the matrix \mathbf{P} describes the q -dimensional null space of the matrix $[\mathbf{N}^T]^T$.

3.1.3.2 Semi-batch reactors

A semi-batch reactor has no outlet ($\omega = 0$). The representation in terms of vessel extents becomes:

$$\dot{x}_{r,i}(t) = r_{v,i}(t), \quad x_{r,i}(0) = 0, \quad i = 1, \dots, R, \quad (3.20a)$$

$$\dot{x}_{in,j}(t) = u_{in,j}(t), \quad x_{in,j}(0) = 0, \quad j = 1, \dots, p, \quad (3.20b)$$

$$\mathbf{x}_{iv}(t) = \mathbf{0}_q. \quad (3.20c)$$

Transformation from \mathbf{x} to \mathbf{n} : The numbers of moles $\mathbf{n}(t)$ can be expressed as:

$$\mathbf{n}(t) = \mathbf{N}^T \mathbf{x}_r(t) + \mathbf{W}_{in} \mathbf{x}_{in}(t) + \mathbf{n}_0. \quad (3.21)$$

Transformation from \mathbf{n} to \mathbf{x} : For a semi-batch reactor,

$$\begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{in}(t) \\ \mathbf{x}_{iv}(t) \end{bmatrix} := \mathcal{T} (\mathbf{n}(t) - \mathbf{n}_0), \quad (3.22)$$

where, $\mathcal{T} = [\mathbf{N}^T \ \mathbf{W}_{in} \ \mathbf{P}]^{-1}$ and the matrix \mathbf{P} describes the q -dimensional null space of the matrix $[\mathbf{N}^T \ \mathbf{W}_{in}]^T$.

3.1.3.3 CSTR

In a CSTR, $u_{out}(t)$ is computed from Eq. (2.7) and $m(t) = V_0 \rho(t)$, with V_0 the constant volume, as follows:

$$u_{out}(t) = \mathbf{1}_p^T \mathbf{u}_{in}(t) - V_0 \dot{\rho}(t). \quad (3.23)$$

- If the density varies, the system (3.10a)-(3.10d) cannot be simplified and thus holds with $q = S - d$.

- If the density is constant, $u_{out}(t) = \mathbf{1}_p^T \mathbf{u}_{in}(t)$ and thus $\omega(t) = \frac{\mathbf{1}_p^T \mathbf{u}_{in}(t)}{m_0}$. In this case, $x_{ic}(t)$ can be computed algebraically from the states $\mathbf{x}_{in}(t)$ as $x_{ic}(t) = 1 - \frac{\mathbf{1}_p^T \mathbf{x}_{in}(t)}{m_0}$, with $m_0 = V_0 \rho$. This can be shown by differentiating the last expression and writing $\dot{\mathbf{x}}_{in}(t)$ and $\dot{x}_{ic}(t)$ using Eqs. (3.10b) and (3.10c). The decoupled system becomes:

$$\dot{x}_{r,i}(t) = r_{v,i}(t) - \omega(t) x_{r,i}(t) \quad x_{r,i}(0) = 0 \quad i = 1, \dots, R \quad (3.24a)$$

$$\dot{x}_{in,j}(t) = u_{in,j}(t) - \omega(t) x_{in,j}(t) \quad x_{in,j}(0) = 0 \quad j = 1, \dots, p \quad (3.24b)$$

$$x_{ic}(t) = 1 - \frac{\mathbf{1}_p^T \mathbf{x}_{in}(t)}{m_0} \quad (3.24c)$$

$$\mathbf{x}_{iv}(t) = \mathbf{0}_q. \quad (3.24d)$$

Hence, the system is of order $R + p$ with $q = S - R - p$. The numbers of moles $\mathbf{n}(t)$ can be expressed as:

$$\mathbf{n}(t) = \mathbf{N}^T \mathbf{x}_r(t) + \left(\mathbf{W}_{in} - \frac{\mathbf{n}_0 \mathbf{1}_p^T}{m_0} \right) \mathbf{x}_{in}(t) + \mathbf{n}_0. \quad (3.25)$$

Example 3.1 (Acetoacetylation of Pyrrole cont'd..)

Consider the acetoacetylation of pyrrole system introduced in the previous example but in a CSTR of constant volume and constant density. For this mode of operation, the representation in terms of vessel extents can be written as:

$$\begin{aligned} \dot{x}_{r,1}(t) &= V(t)r_1(t) - \omega(t)x_{r,1}(t), & x_{r,1}(0) &= 0, \\ \dot{x}_{r,2}(t) &= V(t)r_2(t) - \omega(t)x_{r,2}(t), & x_{r,2}(0) &= 0, \\ \dot{x}_{r,3}(t) &= V(t)r_3(t) - \omega(t)x_{r,3}(t), & x_{r,3}(0) &= 0, \\ \dot{x}_{r,4}(t) &= V(t)r_4(t) - \omega(t)x_{r,4}(t), & x_{r,4}(0) &= 0, \\ \dot{x}_{in}(t) &= u_{in}(t) - \omega(t)x_{in}(t), & x_{in}(0) &= 0, \\ x_{ic}(t) &= 1 - \frac{\mathbf{1}_p^T \mathbf{x}_{in}(t)}{m_0}, \\ x_{iv}(t) &= 0. \end{aligned}$$

3.1.4 Vessel extents representation with heat balance

The representation in terms of vessel extents for a system with heat exchange can be written in terms of five parts, namely, $\mathbf{x}_r(t)$, $x_{ex}(t)$, $\mathbf{x}_{in}(t)$ and $x_{ic}(t)$ that are associated with the reactions, the heat exchange, the inlets and the initial conditions, and $\mathbf{x}_{iv}(t)$ that are invariant:

$$\begin{aligned}
\dot{\mathbf{x}}_r(t) &= \mathbf{r}_v(t) - \omega(t) \mathbf{x}_r(t), & \mathbf{x}_r(0) &= \mathbf{0}_R, \\
\dot{x}_{ex}(t) &= q_{ex}(t) - \omega(t) x_{ex}(t), & x_{ex}(0) &= 0, \\
\dot{\mathbf{x}}_{in}(t) &= \mathbf{u}_{in}(t) - \omega(t) \mathbf{x}_{in}(t), & \mathbf{x}_{in}(0) &= \mathbf{0}_p, \\
\dot{x}_{ic}(t) &= -\omega(t) x_{ic}(t), & x_{ic}(0) &= 1, \\
\mathbf{x}_{iv}(t) &= \mathbf{0}_q,
\end{aligned} \tag{3.26}$$

where x_{ex} is the extent of heat exchange expressed in kJ. Note that the extents \mathbf{x}_r , \mathbf{x}_{in} and x_{ic} in Eq. (3.26) are those in Eqs. (3.10a)-(3.10c), which confirms the fact that the transformed model given in Eqs. (3.10a)-(3.10c) can be used to describe the reactions and flows in a non-isothermal reactor also in the absence of a heat balance.

The numbers of moles $\mathbf{n}(t)$ and the energy $Q(t)$ can be obtained from the transformed variables as follows:

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{n}(t) \\ Q(t) \end{bmatrix} = \mathcal{A} \mathbf{x}_r(t) + \mathbf{b} x_{ex}(t) + \mathcal{C} \mathbf{x}_{in}(t) + \mathbf{z}_0 x_{ic}(t). \tag{3.27}$$

If $\text{rank}([\mathcal{A} \ \mathbf{b} \ \mathcal{C} \ \mathbf{z}_0]) = R + p + 2$, the linear transformation $\mathcal{T} = [\mathcal{A} \ \mathbf{b} \ \mathcal{C} \ \mathbf{z}_0 \ \mathbf{P}]^{-1}$ transforms the representation in terms of $\mathbf{z}(t)$ given in Eq. (2.20) into five parts,

$$\begin{bmatrix} \mathbf{x}_r(t) \\ x_{ex}(t) \\ \mathbf{x}_{in}(t) \\ x_{ic}(t) \\ \mathbf{x}_{iv}(t) \end{bmatrix} = \mathcal{T} \mathbf{z}(t) := \begin{bmatrix} \mathbf{R} \\ \mathbf{h}^\top \\ \mathbf{F} \\ \mathbf{i}^\top \\ \mathbf{P}^+ \end{bmatrix} \mathbf{z}(t). \tag{3.28}$$

The matrix \mathbf{P} describes the q -dimensional null space of the matrix $[\mathcal{A} \ \mathbf{b} \ \mathcal{C} \ \mathbf{z}_0]^\top$, with $q = (S + 1) - (R + p + 2)$.

3.1.5 Use of flowrate information

In this section, the procedure for simplifying the linear transformation from numbers of moles to vessel extents when the flowrates (exogenous inputs) are known is explained for homogeneous reaction systems. Also, a procedure for computing the vessel extent of heat transfer from the measurements in the jacket is explained.

3.1.5.1 From material balance equations

If the inlet and outlet flowrates $\mathbf{u}_{in}(t)$ and $u_{out}(t)$ are known, one can compute $\mathbf{x}_{in}(t)$ and $x_{ic}(t)$ using Eqs. (3.10b) and (3.10c). Then, Eq. (3.11) is used to compute the contribution of the reactions, labeled as the numbers of moles in vessel reaction-variant (vRV) form, as follows:

$$\mathbf{n}^{vRV}(t) := \mathbf{n}(t) - \mathbf{W}_{in} \mathbf{x}_{in}(t) - \mathbf{n}_0 x_{ic}(t), \quad (3.29)$$

which leads to:

$$\mathbf{n}^{vRV}(t) = \mathbf{N}^T \mathbf{x}_r(t), \quad (3.30)$$

or in differential form,

$$\dot{\mathbf{n}}^{vRV}(t) = \mathbf{N}^T \mathbf{r}_v(t) - \omega(t) \mathbf{n}^{vRV}(t), \quad \mathbf{n}^{vRV}(0) = \mathbf{0}_S. \quad (3.31)$$

If $\text{rank}(\mathbf{N}) = R$, the extents of reaction can be computed from Eq. (3.30) as:

$$\mathbf{x}_r(t) = \mathbf{N}^{T+} \mathbf{n}^{vRV}(t) \quad (3.32)$$

Remark 3.7

Since the rank of the matrix \mathbf{N} is always equal to R , the transformation from the numbers of moles in vessel reaction-variant form to the extents of reactions is always possible.

3.1.5.2 From material and heat balance equations

For this case, if the inlet and outlet flowrates are known, one can compute $\mathbf{x}_{in}(t)$ and $x_{ic}(t)$ according to Eqs. (3.10b) and (3.10c). Then, Eq. (3.27) is used to write $\mathbf{z}(t)$ in the vessel reaction and heat-transfer variant ($vRHHV$) form as:

$$\mathbf{z}^{vRHHV}(t) = [\mathcal{A} \ \mathbf{b}] \begin{bmatrix} \mathbf{x}_r(t) \\ x_{ex}(t) \end{bmatrix} = \mathbf{z}(t) - \mathcal{C} \mathbf{x}_{in}(t) - \mathbf{z}_0 x_{ic}(t). \quad (3.33)$$

If $\text{rank}([\mathcal{A} \ \mathbf{b}]) = R + 1$, then the extents of reaction and the extent of heat transfer can be computed from Eq. (3.33) as:

$$\begin{bmatrix} \mathbf{x}_r(t) \\ x_{ex}(t) \end{bmatrix} = [\mathcal{A} \ \mathbf{b}]^+ \mathbf{z}^{vRHHV}(t) \quad (3.34)$$

Note that the extent of heat transfer $x_{ex}(t)$ can also be computed from the heat generated in the jacket $Q_{ja}(t)$. The extent of heat transfer in the jacket $x_{ex,ja}(t)$ can be defined as:

$$\dot{x}_{ex,ja}(t) = -q_{ex}(t) = \dot{Q}_{ja}(t) - \check{q}_{ja,in}(t), \quad x_{ex,ja}(0) = 0. \quad (3.35)$$

If the heat generated in the jacket $Q_{ja}(t)$ is known, then the extent of heat transfer $x_{ex,ja}$ can be computed using:

$$x_{ex,ja}(t) = Q_{ja} - \int_0^t \check{q}_{ja,in}(t) \quad (3.36)$$

By defining a discounting variable δ_h that accounts for the flowrate differences between the reactor and the jacket, the extent of heat transfer $x_{ex}(t)$ can be computed from $x_{ex,ja}(t)$. We define, $x_{ex}(t) = \delta_h(t) - x_{ex,ja}(t)$, where $\delta_h(t)$ is obtained by integrating the following differential equation:

$$\dot{\delta}_h(t) = -\omega(t)(\delta_h(t) - x_{ex,ja}(t)), \quad \delta_h(0) = 0. \quad (3.37)$$

3.2 Heterogeneous Reaction Systems

This section extends the representation in terms of vessel extents obtained for homogeneous reaction systems to heterogeneous fluid-fluid reaction systems.

3.2.1 Vessel extents representation

For the representation of heterogeneous reaction systems in terms of vessel extents, an extent of mass transfer is introduced:

Definition 3.6 (Extent of mass transfer, $x_{m,f,k}$)

The vessel extent of mass transfer of the k th species transferring to phase F is defined as the mass of the k th species entering that phase that is still inside the reactor.

$$\dot{x}_{m,f,k}(t) = \zeta_{f,k}(t) - \omega_f(t) x_{m,f,k}(t), \quad x_{m,f,k}(0) = 0 \quad \forall k = 1, \dots, p_m \quad (3.38)$$

The extent of mass transfer is positive if a species is added to phase F and is negative if it leaves phase F . This follows the convention defined earlier for the sign of ζ_f (see Eq. (2.22)). The representation in terms of vessel extents for phase F is:

$$\dot{\mathbf{x}}_{r,f}(t) = \mathbf{r}_{v,f}(t) - \omega_f(t) \mathbf{x}_{r,f}(t), \quad \mathbf{x}_{r,f}(0) = \mathbf{0}_{R_f}, \quad (3.39a)$$

$$\dot{\mathbf{x}}_{m,f}(t) = \boldsymbol{\zeta}_f(t) - \omega_f(t) \mathbf{x}_{m,f}(t), \quad \mathbf{x}_{m,f}(0) = \mathbf{0}_{p_m}, \quad (3.39b)$$

$$\dot{\mathbf{x}}_{in,f}(t) = \mathbf{u}_{in,f}(t) - \omega_f(t) \mathbf{x}_{in,f}(t), \quad \mathbf{x}_{in,f}(0) = \mathbf{0}_{p_f}, \quad (3.39c)$$

$$\dot{x}_{ic,f}(t) = -\omega_f(t) x_{ic,f}(t), \quad x_{ic,f}(0) = 1, \quad (3.39d)$$

$$\mathbf{x}_{iv,f}(t) = \mathbf{0}_{q_f}. \quad (3.39e)$$

Transformation from \mathbf{x}_f to \mathbf{n}_f : The computation of the numbers of moles $\mathbf{n}_f(t)$ from $\mathbf{x}_f(t)$ is:

$$\mathbf{n}_f(t) = \mathbf{N}_f^T \mathbf{x}_{r,f}(t) + \mathbf{W}_{m,f} \mathbf{x}_{m,f}(t) + \mathbf{W}_{in,f} \mathbf{x}_{in,f}(t) + \mathbf{n}_{f0} x_{ic,f}(t). \quad (3.40)$$

Transformation from \mathbf{n}_f to \mathbf{x}_f For phase F , if $\text{rank} \left([\mathbf{N}_f^T \ \mathbf{W}_{m,f} \ \mathbf{W}_{in,f} \ \mathbf{n}_{f0}] \right) = R_f + p_m + p_f + 1 = d_f$, the linear transformation $\mathcal{T}_f := [\mathbf{N}_f^T \ \mathbf{W}_{m,f} \ \mathbf{W}_{in,f} \ \mathbf{n}_{f0} \ \mathbf{P}_f]^{-1}$

transforms $\mathbf{n}_f(t)$ into five parts, namely, $\mathbf{x}_{r,f}(t)$, $\mathbf{x}_{m,f}(t)$, $\mathbf{x}_{in,f}(t)$ and $x_{ic,f}(t)$ that are associated with the reactions, the mass transfers, the inlets and the initial conditions, and $\mathbf{x}_{iv,f}(t)$ that are invariant:

$$\begin{bmatrix} \mathbf{x}_{r,f}(t) \\ \mathbf{x}_{m,f}(t) \\ \mathbf{x}_{in,f}(t) \\ x_{ic,f}(t) \\ \mathbf{x}_{iv,f}(t) \end{bmatrix} = [\mathbf{N}_f^T \quad \mathbf{W}_{m,f} \quad \mathbf{W}_{in,f} \quad \mathbf{n}_{f0} \quad \mathbf{P}_f]^{-1} \mathbf{n}_f(t), \quad (3.41)$$

where the matrix \mathbf{P}_f of dimension $S_f \times q_f$ describes the q_f -dimensional null space of the matrix $[\mathbf{N}_f^T \quad \mathbf{W}_{m,f} \quad \mathbf{W}_{in,f} \quad \mathbf{n}_{f0}]^T$, with $q_f = S_f - R_f - p_m - p_f - 1$.

3.2.2 Use of flowrate information

If the flowrates $\mathbf{u}_{in,f}(t)$ and $u_{out,f}(t)$ are known, the extents of inlet $\mathbf{x}_{in,f}(t)$ and of initial conditions $x_{ic,f}(t)$ can be computed directly by integrating Eqs. (3.39c) and (3.39d). The numbers of moles can then be written in vessel reaction and mass-transfer variant (vRMV) form as:

$$\mathbf{n}_f^{vRMV}(t) = [\mathbf{N}_f^T \quad \mathbf{W}_{m,f}] \begin{bmatrix} \mathbf{x}_{r,f}(t) \\ \mathbf{x}_{m,f}(t) \end{bmatrix} = \mathbf{n}_f(t) - \mathbf{W}_{in,f} \mathbf{x}_{in,f}(t) - \mathbf{n}_{f0} x_{ic,f}(t). \quad (3.42)$$

The $R_f + p_m$ extents of reaction and of mass transfer can be computed from Eq. (3.42) as:

$$\begin{bmatrix} \mathbf{x}_{r,f}(t) \\ \mathbf{x}_{m,f}(t) \end{bmatrix} = [\mathbf{N}_f^T \quad \mathbf{W}_{m,f}]^+ \mathbf{n}_f^{vRMV}(t). \quad (3.43)$$

Formally, since $\zeta_l = -\zeta_g$, R_l extents of reaction must be computed from phase L and R_g from phase G , whereas the p_m extents of mass transfer can come from either of the two phases; for instance, p_{m_l} extents of mass transfer can be computed from phase L and the remaining p_{m_g} extents from phase G , with $p_{m_l} + p_{m_g} = p_m$. If necessary, the extents of mass transfer obtained from one phase can be converted to extents of mass transfer in the other phase using a discounting variable δ_m , which accounts for the difference between the outlet flows of the two phases. For example, consider that the extents of mass transfer in phase G , $\mathbf{x}_{m,g}(t)$ is known. The corresponding extents of mass transfer in phase L , can be computed as $\mathbf{x}_{m,l}(t) = \delta_m(t) - \mathbf{x}_{m,g}(t)$, with the auxiliary variables $\delta_m(t)$ obtained by integrating the following equation:

$$\dot{\delta}_m(t) = -\omega_l(t)\delta_m(t) + (\omega_l(t) - \omega_g(t))\mathbf{x}_{m,g}(t), \quad \delta_m(0) = 0. \quad (3.44)$$

Example 3.2 (Chlorination of Butanoic Acid)

Consider the chlorination of butanoic acid (BA) given in Example 2.4. Assume that the reaction system is operated in a semi-batch reactor, with Cl_2 fed in the gas phase and BA present initially in the liquid phase. For phase L the representation in terms of vessel extents is:

$$\begin{aligned} \dot{x}_{r,l,1}(t) &= V_l(t) r_{l,1}(t), & x_{r,l,1}(0) &= 0, \\ \dot{x}_{r,l,2}(t) &= V_l(t) r_{l,2}(t), & x_{r,l,2}(0) &= 0, \\ \dot{x}_{m,\text{Cl}_2,l}(t) &= \zeta_{\text{Cl}_2,l}(t), & x_{m,\text{Cl}_2,l}(0) &= 0, \\ \dot{x}_{m,\text{HCl},l}(t) &= \zeta_{\text{HCl},l}(t), & x_{m,\text{HCl},l}(0) &= 0. \end{aligned}$$

The numbers of moles in phase L can be reconstructed using Eq. (3.40). Since phase L has five species, $d_l = R_l + p_m = 4$, the $\text{rank}([\mathbf{N}_l^T \mathbf{W}_{m,l}]) = d_l$ and the transformation \mathcal{T}_l can be applied to transform the numbers of moles to vessel extents.

For phase G , the representation in terms of vessel extents is:

$$\begin{aligned} \dot{x}_{m,\text{Cl}_2,g}(t) &= \zeta_{\text{Cl}_2,g}(t), & x_{m,\text{Cl}_2,g}(0) &= 0, \\ \dot{x}_{m,\text{HCl},g}(t) &= \zeta_{\text{HCl},g}(t), & x_{m,\text{HCl},g}(0) &= 0, \\ \dot{x}_{in,g}(t) &= u_{in,g}(t), & x_{in,g}(0) &= 0. \end{aligned}$$

For phase G has two species, $d_g = p_m + p_g = 3$, and $\text{rank}([\mathbf{W}_{m,g} \mathbf{W}_{in,g}]) = 2 < d_g$ and hence the transformation \mathcal{T}_g cannot be used to transform \mathbf{n}_g to \mathbf{x}_g . However, it is possible to go from \mathbf{x}_g to \mathbf{n}_g using Eq. (3.40).

3.3 Simulated Examples

In this section, the material balances and the vessel-extents representations of homogeneous and heterogeneous reaction systems are illustrated via the simulated examples of a homogeneous and of a heterogeneous reaction system.

3.3.1 Homogeneous reaction system

Consider the acetoacetylation of pyrrole system given in Example 2.3. The reaction system is operated in a reactor with one inlet and one outlet streams. The inlet stream contains the species A, B and K, with the composition matrix given by \mathbf{W}_{in} in Eq. (3.16) and a constant inlet flowrate 2 g min^{-1} . The reactor also has an outlet stream with a constant flowrate 2 g min^{-1} . The reactor initially contains 2 mol of A, 5 mol

of B and 0.5 mol of catalyst K, for an initial volume of 0.593 L. Assuming constant density, the volume remains constant at 0.593 L. The following rate models for each of the four reactions are used for simulation:

$$r_1 = k_1 c_A c_B c_K \quad (3.45a)$$

$$r_2 = k_2 c_B^2 c_K \quad (3.45b)$$

$$r_3 = k_3 c_B \quad (3.45c)$$

$$r_4 = k_4 c_C c_B c_K. \quad (3.45d)$$

The values of the rate constants are $k_1 = 0.0530 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$, $k_2 = 0.1280 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$, $k_3 = 0.0280 \text{ min}^{-1}$ and $k_4 = 0.003 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$. The reaction system is simulated using the material balance equations given in Eq. (2.6). The simulated numbers of moles of all the seven species are shown in Fig. 3.2.

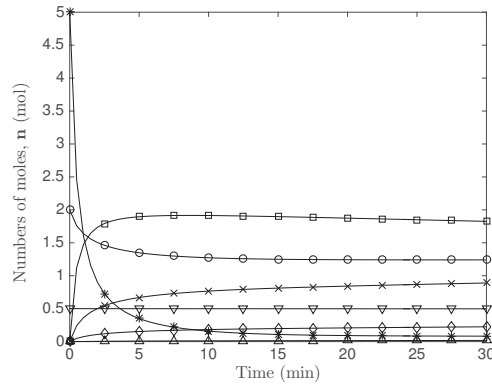
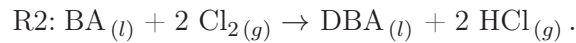
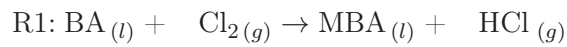


Figure 3.2 Acetoacetylation of pyrrole: Simulated numbers of moles of species A (o), B (*), C (×), D (□), E (◇), F (△) and K (▽).

The reaction system is also simulated using the representation in terms of vessel extents given in Eqs. (3.10a) - (3.10c). The simulated vessel extents of reaction, inlet and initial conditions are shown in Fig. 3.3.

3.3.2 Heterogeneous reaction system

Consider the chlorination of butanoic acid system explained in Example 2.4. The reaction can be represented by the following scheme:



The rate laws for reactions R1 and R2 and the mass transfer rates of Cl_2 and HCl are:

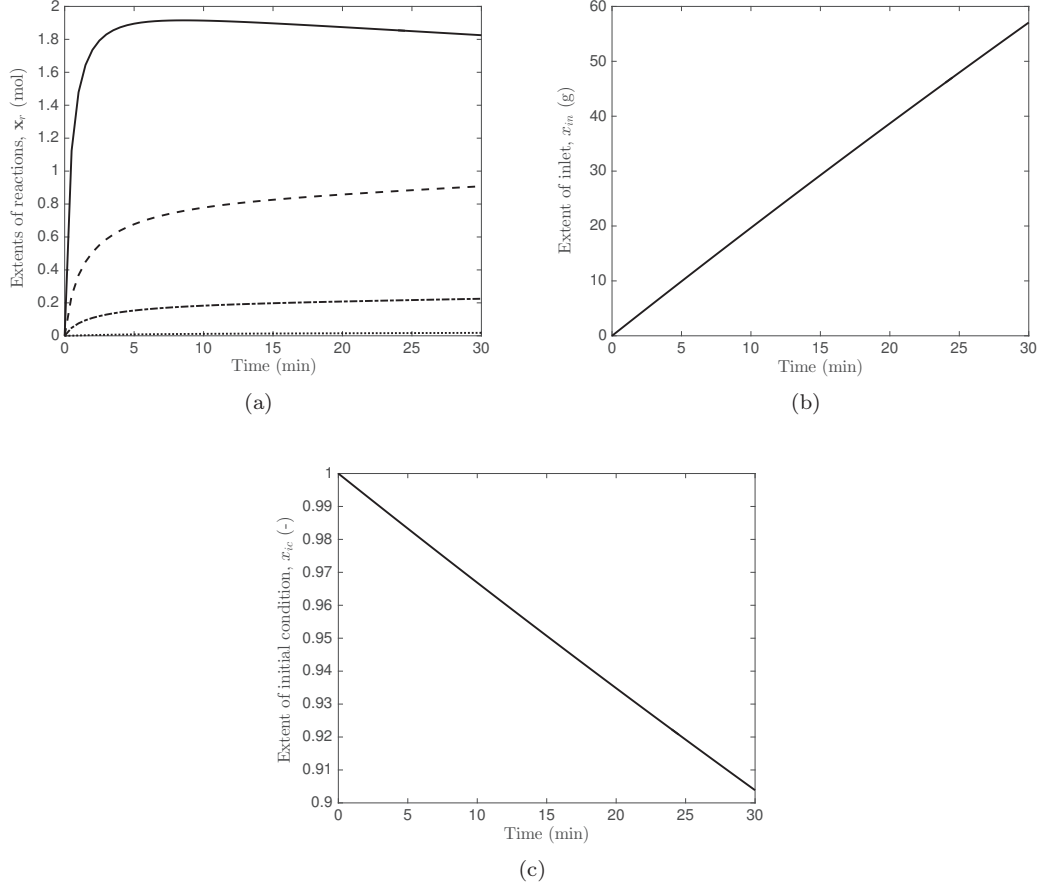


Figure 3.3 Acetoacetylation of pyrrole: (a) Simulated vessel extents of reaction $x_{r,1}$ (—), $x_{r,2}$ (---), $x_{r,3}$ (-.-) and $x_{r,4}$ (···), (b) simulated vessel extents of inlet (\mathbf{x}_{in}), and (c) simulated vessel extents of initial conditions (x_{ic}).

$$r_1 = \frac{k_1 \sqrt{c_{MBA} + k_2}}{1 + k_3 c_{Cl_2}} \left(\frac{c_{BA}}{c_{BA} + \epsilon_1} \right) \left(\frac{c_{Cl_2}}{c_{Cl_2} + \epsilon_2} \right) \quad (3.46)$$

$$r_2 = k_4 r_1 c_{Cl_2} \quad (3.47)$$

$$\zeta_{Cl_2,g} = k_{Cl_2} A V_g M_{w,Cl_2} (c_{Cl_2} - c_{Cl_2}^*) \quad (3.48)$$

$$\zeta_{HCl,l} = k_{HCl} A V_l M_{w,HCl} (c_{HCl} - c_{HCl}^*), \quad (3.49)$$

where k_1 , k_2 , k_3 and k_4 are rate constants, $c_{Cl_2}^*$ and c_{HCl}^* are the equilibrium concentrations of Cl_2 and HCl at the interface, calculated using Henry constants, k_{Cl_2} and k_{HCl} are the molar transfer coefficients of Cl_2 and HCl and A is the specific interfacial area. The values of the reaction parameters taken from [14] are $k_1 = 0.0044 \text{ (kmol m}^{-3}\text{)}^{1/2} \text{ s}^{-1}$, $k_2 = 0.0088 \text{ kmol m}^{-3} \text{ s}^{-1}$, $k_3 = 1.3577 \text{ m}^3 \text{ kmol}^{-1}$, $k_4 = 0.1$, $A = 254.9 \text{ m}^{-1}$, $k_{Cl_2} = 0.666 \times 10^{-4} \text{ m s}^{-1}$ and $k_{HCl} = 0.845 \times 10^{-4} \text{ m s}^{-1}$ at 403K.

For the simulation, 10 kmol of butanoic acid, a small amount of MBA and 100 kmol of solvent are initially loaded in the reactor. The reactor is operated in semi-batch mode and continuously fed with chlorine gas with a mass flowrate of $u_{in,g} = 972 \text{ kg h}^{-1}$. The reactor volume V_r is constant at 9 m^3 , with the gas and liquid phases initially occupying 1.844 m^3 and 7.156 m^3 , respectively. The simulated numbers of moles of all the species in the liquid and the gas phases are shown in Fig. 3.4.

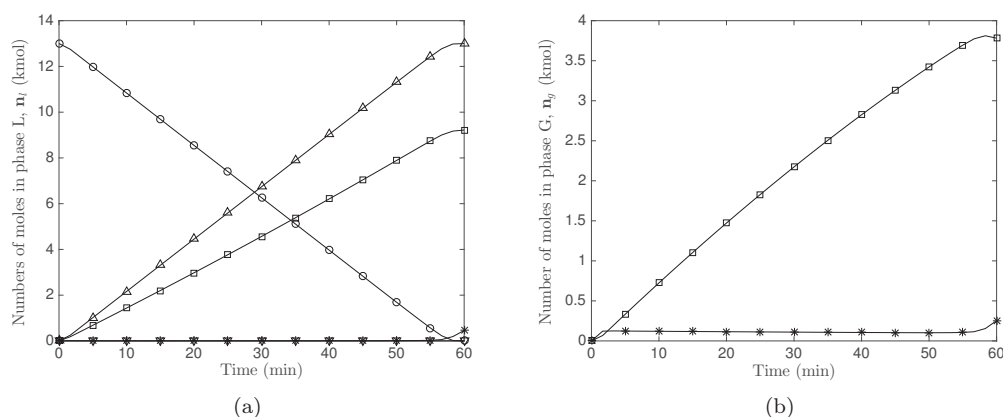


Figure 3.4 Chlorination of butanoic acid: (a) Simulated numbers of moles of liquid species BA (o), Cl_2 (*), MBA (Δ), HCl (\square) and DBA (∇), and (b) simulated numbers of moles of gaseous species Cl_2 (*) and HCl (\square).

The phases L and G are also simulated using the representation in terms of vessel extents. The simulated vessel extents of reaction, mass transfer, and inlet are shown in Figs. 3.5 and 3.6.

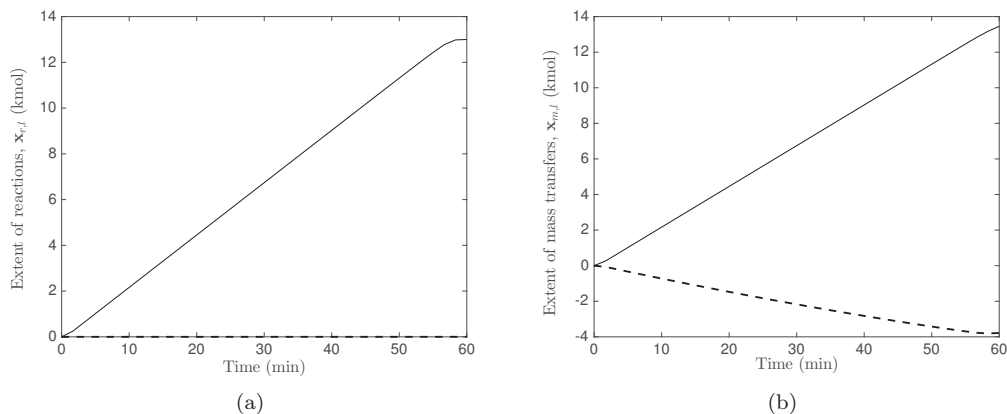


Figure 3.5 Chlorination of butanoic acid: (a) Simulated vessel extents of reaction in liquid phase $x_{r,1}$ (—) and $x_{r,2}$ (- -), and (b) simulated vessel extents of mass transfer in liquid phase x_{m,Cl_2} (—) and $x_{m,HCl}$ (- -).

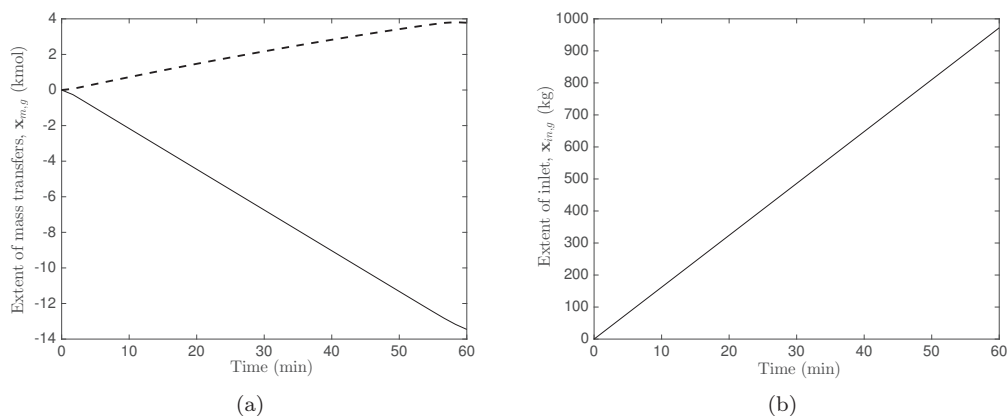


Figure 3.6 Chlorination of butanoic acid: (a) Simulated vessel extents of mass transfer in gas phase x_{m,Cl_2} (—) and $x_{m,HCl}$ (- -), and (b) simulated vessel extents of inlet in gas phase $x_{in,g}$.

3.4 Summary

This chapter has introduced an alternative representation for homogeneous chemical reaction systems in terms of vessel extents. For homogeneous reaction systems, this representation has R vessel extents of reaction, p extents of inlet and extent of initial conditions. Each vessel extent of reaction represents the amount in numbers of moles produced or consumed by that reaction that is still inside the reactor, while a vessel extent of inlet represents the mass added by an inlet stream that is still inside the reactor. The extent of initial conditions represents the fraction of initial conditions that is still inside the reactor. This chapter has also introduced a linear transformation that

transforms the representation in terms of numbers of moles for a homogeneous reaction systems to the representation in terms of vessel extents. This chapter has introduced an additional vessel extent, namely, the vessel extent of heat exchange, to capture the heat transfer between the reactor and its jacket. Additionally, a procedure to simplify the linear transformation from numbers of moles to vessel extents has been presented in the case the flowrates are available.

The vessel extent representation has been extended to heterogeneous reaction systems. In this case, a phase F in the reaction system has R_f vessel extents of reaction, p_f extents of inlet, p_m extents of mass transfer and one extent of initial conditions. The procedure for transforming the representation in terms of numbers of moles to vessel extents and vice versa has also been presented.

Chapter 4

Data Reconciliation

Chapter 3 introduced the representation of chemical reaction systems in terms of vessel extents. In practice, in the absence of a kinetic model, the vessel extents have to be inferred from the measured (noisy) numbers of moles and cannot be measured directly. In this chapter, we aim to formulate a data reconciliation (DR) problem to improve the accuracy of the numbers of moles and vessel extents. First, the DR problem is formulated in terms of the numbers of moles and then in terms of vessel extents.

This chapter shows the existence of knowledge-based constraints, namely, constraints that are satisfied at all times such as monotonicity and curvature (concavity/convexity) which provides some dynamic information in the absence of a kinetic model. The nature and type of these constraints depend on the operating conditions of the reactor (batch, semi-batch or open with inlets and outlet). It is shown that the state representation in terms of extents helps (i) simplify the identification of knowledge-based constraints, and (ii) impose additional shape constraints compared to a representation with classical states. Conditions of validity of these shape constraints are derived. In addition, for cases where it is difficult or even impossible to predict shapes beforehand, we propose a method of shape identification based on measurements. In the presence of these shape constraints, DR solves a weighted least-squares problem that accounts for all the measurements up to the current time instant (for online applications) or up to the final time (for offline applications).

This chapter is organized as follows. Section 4.1 introduces a systematic procedure for deriving shape constraints both in terms of numbers of moles and vessel extents for a homogeneous reaction systems. Section 4.2 derives the corresponding constraints for a heterogeneous reaction systems. The data-reconciliation problems for homogeneous and heterogeneous reaction systems, in the absence of a kinetic model, are formulated in Section 4.3. Finally, the proposed reconciliation procedure is illustrated via the simulated examples of homogeneous and heterogeneous reaction systems in Section 4.4.

4.1 Homogeneous Reaction Systems

Let $\tilde{\mathbf{n}}(t_h) = \mathbf{n}_{true}(t_h) + \epsilon_n$ denote the S -dimensional vector of noisy numbers of moles measured at the time instant t_h , $t_h \in [t_1, t_H]$, where t_1 corresponds to the initial

conditions $t = 0$, ϵ_n is an S -dimensional vector of zero-mean Gaussian noise with a constant variance-covariance matrix Σ_n .

4.1.1 Shape constraints

In this subsection, knowledge-based constraints are formulated in terms of both numbers of moles and extents. Moreover, to generate additional constraints, or when shape constraints are not known *a priori*, a methodology is proposed to identify shape constraints from measurements.

4.1.1.1 Knowledge-based constraints

Knowledge-based constraints are those that are known from prior knowledge and are valid for all times. These shape constraints, abstractly denoted \mathcal{K} , can be formulated in terms of the numbers of moles \mathbf{n} or extents \mathbf{x} :

$$\mathcal{K}_n(\mathbf{n}(t)) \leq \mathbf{0}_{k_n} \quad \text{or} \quad \mathcal{K}_x(\mathbf{x}(t)) \leq \mathbf{0}_{k_x}, \quad (4.1)$$

with k_n and k_x the numbers of knowledge-based constraints expressed in terms of numbers of moles and extents, respectively.

For the purpose of attributing shape properties to certain species, we need to introduce very specific classes of species. The two subsets of interest are the subset of non-added non-produced species \mathcal{S}_{np} and the subset of non-removed non-consumed species \mathcal{S}_{nc} . These subsets are defined next for the case of homogeneous reaction systems.

Definition 4.1 (Non-added non-produced species, \mathcal{S}_{np})

A species belongs to the set of non-added non-produced species \mathcal{S}_{np} if the corresponding column of the stoichiometric matrix contains only nonpositive elements and the species is not fed to the reactor via an inlet stream.

Definition 4.2 (Non-removed non-consumed species, \mathcal{S}_{nc})

A species belongs to the set of non-removed non-consumed species \mathcal{S}_{nc} if the corresponding column of the stoichiometric matrix contains only nonnegative elements and the species is not removed from the reactor via an outlet stream.

Remark 4.1

The set \mathcal{S}_{nc} is empty for reactors with an outlet, since all species are removed via the outlet flow.

Knowledge-based constraints in terms of numbers of moles

The following shape constraints in terms of numbers of moles can be defined *a priori*.

Lemma 1: (Non-added non-produced species involved in irreversible reactions)

The number of moles of a non-added non-produced species that is only involved in irreversible reactions is nonincreasing.

Proof: The material balance equation for the non-added non-produced species s can be written as:

$$\dot{n}_s(t) = (\mathbf{N})_s^T \mathbf{r}_v(t) - \omega(t) n_s(t), \quad s \in \mathcal{S}_{np} \quad (4.2)$$

with n_s the number of moles of the species s and $(\mathbf{N})_s^T$ the row corresponding to the species s in \mathbf{N}^T . Since the rates of all irreversible reactions are nonnegative and the stoichiometric coefficients of the species s are all nonpositive, the term $(\mathbf{N})_s^T \mathbf{r}_v(t)$ is nonpositive for all times. Since the term $-\omega(t) n_s(t)$ is also nonpositive, one concludes that $\dot{n}_s(t)$ is nonpositive for all times, and thus the number of moles of the species s is nonincreasing. \square

Lemma 2: (Non-removed non-consumed species involved in irreversible reactions)

The number of moles of a non-removed non-consumed species that is only involved in irreversible reactions is nondecreasing.

Proof: The material balance equation for the non-removed non-consumed species s can be written as:

$$\dot{n}_s(t) = (\mathbf{N})_s^T \mathbf{r}_v(t) + (\mathbf{W}_{in})_s \mathbf{u}_{in}(t), \quad s \in \mathcal{S}_{nc} \quad (4.3)$$

Since the rates of irreversible reactions are nonnegative and the stoichiometric coefficients of the product species s are all nonnegative, the term $(\mathbf{N})_s^T \mathbf{r}_v(t)$ is nonnegative for all times. Since the term $(\mathbf{W}_{in})_s \mathbf{u}_{in}(t)$ is also nonnegative, one concludes that $\dot{n}_s(t)$ is nonnegative for all times, and thus the number of moles of the species s is nondecreasing. \square

Corollary: If a species belongs to both sets \mathcal{S}_{np} and \mathcal{S}_{nc} , namely, it is neither added nor removed and neither consumed nor produced, its number of moles remains constant. This is for example the case of a catalyst.

Note that it is difficult to predict *a priori* the curvature properties of the numbers of moles, as the numbers of moles can be functions of multiple reactions, unlike vessel extents. Furthermore, knowledge-based constraints in terms of numbers of moles are very restrictive since several conditions must be satisfied in Lemmas 1 and 2. Also, *a priori* constraints cannot be imposed to species involved in reversible reactions since their shape can potentially vary during the course of the reaction. However, additional constraints can be inferred from measurements using the procedure described in Section 4.1.1.2

Knowledge-based constraints in terms of extents

Knowledge-based constraints in terms of extents are described for reactors both without and with an outlet.

Reactors without outlet

In the absence of an outlet flow, the system representation of Eq. (3.10) reduces to:

$$\begin{aligned}\dot{\mathbf{x}}_r(t) &= \mathbf{r}_v(t), & \mathbf{x}_r(0) &= \mathbf{0}_R, \\ \dot{\mathbf{x}}_{in}(t) &= \mathbf{u}_{in}(t), & \mathbf{x}_{in}(0) &= \mathbf{0}_p,\end{aligned}\tag{4.4}$$

with $x_{ic} = 1$ and $\mathbf{x}_{iv} = \mathbf{0}_q$.

The following shape constraints are always valid:

Lemma 3: (Extents of inlet)

The extents of inlet $\mathbf{x}_{in}(t)$ are (i) nonnegative nondecreasing functions, and (ii) convex (concave) functions if the corresponding inlet flowrates are nonnegative nondecreasing (nonincreasing)¹.

Proof: The proof of (i) follows from

$$\mathbf{x}_{in}(t_h) = \mathbf{x}_{in}(t_{h-1}) + \int_{t_{h-1}}^{t_h} \mathbf{u}_{in}(t) dt, \quad \forall h = 2, \dots, H, \tag{4.5}$$

and the fact that $\mathbf{u}_{in}(t) \geq \mathbf{0}_p$ for $t \in [t_1, t_H]$.

To prove convexity in (ii), consider the three time instants $t_h < t_{h+1} < t_{h+2}$ in the interval $[t_1, t_H]$. From (4.5) and the fact that the inlet flowrates are nonnegative nondecreasing, it follows that

$$\frac{\mathbf{x}_{in}(t_{h+2}) - \mathbf{x}_{in}(t_{h+1})}{t_{h+2} - t_{h+1}} \geq \mathbf{u}_{in}(t_{h+1}) \geq \frac{\mathbf{x}_{in}(t_{h+1}) - \mathbf{x}_{in}(t_h)}{t_{h+1} - t_h},$$

which proves the convexity property of $\mathbf{x}_{in}(t)$ via the relation that exists between the left and the right terms [20]. Similar arguments can be used to prove concavity. \square

Lemma 4: (Extents of irreversible reactions)

The extents of irreversible reactions are (i) nonnegative nondecreasing functions, and (ii) concave (convex) functions if the corresponding reaction rates are nonnegative nonincreasing (nondecreasing).

Proof: The proof of (i) follows from

¹ In practice, it is required to know the monotonic behaviour of the inlet flowrates but not their exact numerical values.

$$\mathbf{x}_r(t_h) = \mathbf{x}_r(t_{h-1}) + \int_{t_{h-1}}^{t_h} \mathbf{r}_v(t) dt \quad \forall h = 2, \dots, H, \quad (4.6)$$

and the fact that $\mathbf{r}_v(t) \geq \mathbf{0}_R$ for $t \in [t_1, t_H]$.

To prove concavity in (ii), consider the three time instants $t_h < t_{h+1} < t_{h+2}$ in the interval $[t_1, t_H]$. From (4.6) and the fact that the reaction rates are nonnegative nonincreasing, it follows that

$$\frac{\mathbf{x}_r(t_{h+2}) - \mathbf{x}_r(t_{h+1})}{t_{h+2} - t_{h+1}} \leq \mathbf{r}_v(t_{h+1}) \leq \frac{\mathbf{x}_r(t_{h+1}) - \mathbf{x}_r(t_h)}{t_{h+1} - t_h},$$

which proves the concavity property of $\mathbf{x}_r(t)$ via the relation that exists between the left and the right terms. Similar arguments can be used to prove convexity. \square

Reactors with outlet

For reactors with an outlet flow, shape constraints exist for the extent of initial conditions, and so-called generalized shape constraints can be derived from their differential expressions for the extents of reaction and inlet.

Lemma 5: (Extent of initial conditions)

The extent of initial conditions is a nonnegative nonincreasing function.

Proof: The solution to Eq. (3.10c) is $x_{ic}(t) = e^{-\int_0^t \omega(t) dt}$. It follows that $x_{ic}(t)$ cannot be negative and, from $\omega(t) \geq 0$, that it is nonnegative nonincreasing. \square

In the presence of an outlet stream, the extents of reaction and the extents of inlet are described by Eqs. (3.10a) and (3.10b), respectively. Lemmas 3 and 4 no longer hold due to the presence of the discounting terms $\omega(t)\mathbf{x}_r(t)$ for the extents of reaction and $\omega(t)\mathbf{x}_{in}(t)$ for the extents of inlet. Nevertheless, one can derive shape constraints from their differential expressions, as illustrated next.

Proposition 4.1 (Generalized shape constraints for the extents of reaction and inlet)

The generalized shape constraints for the extents of irreversible reactions and the extents of inlet are as follows:

$$\dot{\mathbf{x}}_r(t) + \omega(t) \mathbf{x}_r(t) = \mathbf{r}_v(t) \geq \mathbf{0}_R \quad (4.7a)$$

$$\dot{\mathbf{x}}_{in}(t) + \omega(t) \mathbf{x}_{in}(t) = \mathbf{u}_{in}(t) \geq \mathbf{0}_p. \quad (4.7b)$$

Proof: Eqs. (4.7a) and (4.7b) are obtained by re-arranging Eqs. (3.10a) and (3.10b) and using the fact that $\mathbf{r}_v(t) \geq \mathbf{0}_R$ (for irreversible reactions) and $\mathbf{u}_{in}(t) \geq \mathbf{0}_p$.

If measurements of the outlet flowrate and of the mass are available, the values $\omega(t_h) = \frac{u_{out}(t_h)}{m(t_h)}$, $\forall h = 1, 2, \dots, H$, are known, and Eqs. (4.7a) and (4.7b) can be evaluated by discretization of the derivative terms at the sampling instants:

$$\frac{\mathbf{x}_r(t_h) - \mathbf{x}_r(t_{h-1})}{t_h - t_{h-1}} + \omega(t_{h-1}) \mathbf{x}_r(t_{h-1}) \geq \mathbf{0}_R, \quad (4.8)$$

$$\frac{\mathbf{x}_{in}(t_h) - \mathbf{x}_{in}(t_{h-1})}{t_h - t_{h-1}} + \omega(t_{h-1}) \mathbf{x}_{in}(t_{h-1}) \geq \mathbf{0}_p, \quad \forall h = 2, \dots, H. \quad (4.9)$$

On the other hand, if measurements of the outlet flowrate and the mass are not available, Eq. (3.10c) can be discretized and re-arranged to obtain an expression for $\omega(t)$ in terms of the extent of initial conditions:

$$\omega(t_{h-1}) = \frac{x_{ic}(t_{h-1}) - x_{ic}(t_h)}{(t_h - t_{h-1}) x_{ic}(t_{h-1})}, \quad \forall h = 2, \dots, H. \quad (4.10)$$

Replacing $\omega(t_{h-1})$ in Eqs. (4.8) and (4.9) by its value in Eq. (4.10) gives the discretized shape constraints:

$$\mathbf{x}_r(t_h) - \frac{x_{ic}(t_h)}{x_{ic}(t_{h-1})} \mathbf{x}_r(t_{h-1}) \geq \mathbf{0}_R, \quad (4.11a)$$

$$\mathbf{x}_{in}(t_h) - \frac{x_{ic}(t_h)}{x_{ic}(t_{h-1})} \mathbf{x}_{in}(t_{h-1}) \geq \mathbf{0}_p, \quad \forall h = 2, \dots, H. \quad (4.11b)$$

4.1.1.2 Measurement-based constraints

Knowledge-based constraints cannot always be guaranteed *a priori*. Nevertheless, some constraints could well be present in the measurements.

For example, it is well known that a species that is formed and consumed by different reactions with comparable reaction rates in batch conditions will exhibit an increase followed by a decrease. Since such a species is neither part of \mathcal{S}_{np} nor of \mathcal{S}_{nc} – it is at the same time both a reactant and a product – a knowledge-based constraint cannot be defined for this type of behavior.

This section introduces a procedure to identify valid shape constraints for a species or an extent based on measurements. For the sake of simplicity, the procedure is described for the i th extent, but the same procedure also applies to an individual number of moles.

Proposition 4.2 (Procedure to identify measurement-based constraints)

The procedure to identify the shape constraints from the measured numbers of moles $\tilde{\mathbf{n}}$ is as follows:

1. Estimate the extents $\hat{x}_i(t_h)$, $\forall h = 1, \dots, H$, by using either the linear transformation given by Eq. (3.14) or solution to the DR problem described in Section 4.3 in the presence of positivity constraints and additional knowledge-based constraints (if any). The corresponding standard deviation $\hat{\sigma}_{x_i}(t_h)$ is also estimated.

2. Estimate the first and second derivatives of $\hat{x}_i(t_h)$, denoted $\dot{\hat{x}}_i(t_h)$ and $\ddot{\hat{x}}_i(t_h)$, as well as their upper and lower bounds $\dot{\hat{x}}_i^L(t_h)$, $\dot{\hat{x}}_i^U(t_h)$, $\ddot{\hat{x}}_i^L(t_h)$ and $\ddot{\hat{x}}_i^U(t_h)$. In this work, the procedure suggested by Bunin et al. [24] is used.
3. Identify curvature constraints based on the sign of the upper and lower bounds of the second derivative using time windows (of variable size k) in which the following conditions are satisfied:
 - a. If $\ddot{\hat{x}}_i$ is greater than a positive threshold ($\ddot{\hat{x}}_i^L > 0$) for all points in the time window \mathcal{T}^k , the i th extent is convex in that time window.
 - b. If $\ddot{\hat{x}}_i$ is lower than a negative threshold ($\ddot{\hat{x}}_i^U < 0$) for all points in the time window \mathcal{T}^k , the i th extent is concave in that time window.
 - c. If none of the conditions (a) or (b) are satisfied for the time window \mathcal{T}^k , the i th extent follows no curvature constraint in that window.

Repeat Step 3 for all the time windows in which the i th extent can be convex or concave.

4. Identify monotonicity constraints based on the sign of the upper and lower bounds of the first derivative using time windows (of variable size k) in which the following conditions are satisfied:
 - a. If $\dot{\hat{x}}_i$ is greater than a positive threshold ($\dot{\hat{x}}_i^L > 0$) for all points in the time window \mathcal{T}^k , the i th extent is monotonically increasing in that time window.
 - b. If $\dot{\hat{x}}_i$ is lower than a negative threshold ($\dot{\hat{x}}_i^U < 0$) for all points in the time window \mathcal{T}^k , the i th extent is monotonically decreasing in that time window.
 - c. If none of the conditions (a) or (b) are satisfied for the time window \mathcal{T}^k , the i th extent follows no monotonicity constraint in that window.

Repeat Step 4 for all the time windows in which the i th extent can be monotonically increasing or decreasing.

The constraints identified from measurements are formulated in an abstract way as:

$$\mathcal{M}_n(\mathbf{n}(t)) \leq \mathbf{0}_{m_n} \quad \text{or} \quad \mathcal{M}_x(\mathbf{x}(t)) \leq \mathbf{0}_{m_x}, \quad (4.12)$$

with m_n and m_x the numbers of measurement-based constraints expressed in terms of numbers of moles and extents, respectively.

4.2 Heterogeneous Reaction Systems

In this section, we derive the shape constraints for heterogeneous reaction systems. Let $\tilde{\mathbf{n}}_f(t_h) = \mathbf{n}_{true,f}(t_h) + \epsilon_{n_f}$ denote the S_f -dimensional vector of noisy numbers of moles in phase F measured at the time instant t_h , $t_h \in [t_1, t_H]$, where ϵ_{n_f} is an S_f -dimensional vector of zero-mean Gaussian noise with a constant variance-covariance matrix Σ_{n_f} .

4.2.1 Shape constraints

Similarly to the homogeneous reaction systems, knowledge-based constraints for a heterogeneous reaction system can be derived both in terms of numbers of moles or vessel extents as shown in the next subsection. The procedure for identifying the measurement-based constraints is exactly the same as the procedure in Section 4.1.1.2.

4.2.1.1 Knowledge-based constraints

The knowledge-based shape constraints for heterogeneous reaction systems are abstractly denoted \mathcal{K} and can be formulated in terms of the numbers of moles \mathbf{n}_f or the extents \mathbf{x}_f :

$$\mathcal{K}_{n_f}(\mathbf{n}_f(t)) \leq \mathbf{0}_{k_{n_f}} \quad \text{or} \quad \mathcal{K}_{x_f}(\mathbf{x}_f(t)) \leq \mathbf{0}_{k_{x_f}}, \quad (4.13)$$

with k_{n_f} and k_{x_f} the numbers of knowledge-based constraints expressed in terms of numbers of moles and extents, respectively.

For heterogeneous reaction systems, we introduce two subsets of interest, namely the subset of non-added non-produced species $\mathcal{S}_{np,f}$ and the subset of non-removed non-consumed species $\mathcal{S}_{nc,f}$. These subsets are defined next for the case of heterogeneous reaction systems.

Definition 4.3 (Non-added non-produced species, $\mathcal{S}_{np,f}$)

A species belongs to the set of non-added non-produced species $\mathcal{S}_{np,f}$ if the corresponding column of the stoichiometric matrix contains only nonpositive elements and the species is not added to phase F via an inlet stream or mass transfer.

Definition 4.4 (Non-removed non-consumed species, $\mathcal{S}_{nc,f}$)

A species belongs to the set of non-removed non-consumed species $\mathcal{S}_{nc,f}$ if the corresponding column of the stoichiometric matrix contains only nonnegative elements and the species is not removed from phase F via an outlet stream or mass transfer.

4.2.1.2 Knowledge-based constraints in terms of numbers of moles

The following shape constraints in terms of numbers of moles can be defined *a priori*.

Lemma 6: (Non-added non-produced species involved in irreversible reactions)

The number of moles of a non-added non-produced species that is only involved in irreversible reactions is nonincreasing.

Proof: The material balance equation for the non-added non-produced species s can be written as:

$$\dot{n}_{f,s}(t) = (\mathbf{N}_f)_s^T \mathbf{r}_{v,f}(t) + (\mathbf{W}_{m,f})_s \boldsymbol{\zeta}_f(t) - \omega_f(t) n_{f,s}(t), \quad s \in \mathcal{S}_{np,f} \quad (4.14)$$

with $n_{f,s}$ the number of moles of the species s , $(\mathbf{N}_f)_s^T$ the row corresponding to the species s in \mathbf{N}_f^T , and $(\mathbf{W}_{m,f})_s$ the row corresponding to the species s in $\mathbf{W}_{m,f}$.

Since the rates of all irreversible reactions are nonnegative and the stoichiometric coefficients of the species s are all nonpositive, the term $(\mathbf{N}_f)_s^T \mathbf{r}_{v,f}(t)$ is nonpositive for all times. In addition, as the vector $(\mathbf{W}_{m,f})_s$ only contains ones and zeros by construction and the rates of the mass transfer involving the species s are all nonpositive, the term $(\mathbf{W}_{m,f})_s \boldsymbol{\zeta}_f(t)$ is nonpositive for all times. Finally, noting that the term $-\omega_f(t) n_{f,s}(t)$ is also nonpositive, one concludes that $\dot{n}_{f,s}(t)$ is nonpositive for all times, and thus the number of moles of the species s is nonincreasing. \square

Lemma 7: (Non-removed non-consumed species involved in irreversible reactions)

The number of moles of a non-removed non-consumed species that is only involved in irreversible reactions is nondecreasing.

Proof: The material balance equation for the non-removed non-consumed species s can be written as:

$$\dot{n}_{f,s}(t) = (\mathbf{N}_f)_s^T \mathbf{r}_{v,f}(t) + (\mathbf{W}_{m,f})_s \boldsymbol{\zeta}_f(t) + (\mathbf{W}_{in,f})_s \mathbf{u}_{in,f}(t), \quad s \in \mathcal{S}_{nc,f} \quad (4.15)$$

with $(\mathbf{W}_{in,f})_s$ the row corresponding to the species s in $\mathbf{W}_{in,f}$.

Since the rates of all irreversible reactions are nonnegative and the stoichiometric coefficients of the product species s are all nonnegative, the term $(\mathbf{N}_f)_s^T \mathbf{r}_{v,f}(t)$ is nonnegative for all times. In addition, as the vector $(\mathbf{W}_{m,f})_s$ only contains ones and zeros by construction and the rates of the mass transfer involving the species s are all nonnegative, the term $(\mathbf{W}_{m,f})_s \boldsymbol{\zeta}_f(t)$ is nonnegative for all times. Finally, noting that the term $(\mathbf{W}_{in,f})_s \mathbf{u}_{in,f}(t)$ is nonnegative, one concludes that $\dot{n}_{f,s}(t)$ is nonnegative for all times, and thus the number of moles of the species s is nondecreasing. \square

Knowledge-based constraints in terms of extents

Knowledge-based constraints in terms of extents are described for reactors both without and with an outlet.

Reactors without outlet

In the absence of an outlet, the representation in terms of vessel extents for a heterogeneous reaction system given in Eq. (3.39) reduces to:

$$\dot{\mathbf{x}}_{r,f}(t) = \mathbf{r}_{v,f}(t), \quad \mathbf{x}_{r,f}(0) = \mathbf{0}_{R_f}, \quad (4.16a)$$

$$\dot{\mathbf{x}}_{m,f}(t) = \boldsymbol{\zeta}_f(t), \quad \mathbf{x}_{m,f}(0) = \mathbf{0}_{p_m}, \quad (4.16b)$$

$$\dot{\mathbf{x}}_{in,f}(t) = \mathbf{u}_{in,f}(t), \quad \mathbf{x}_{in,f}(0) = \mathbf{0}_{p_f}, \quad (4.16c)$$

$$x_{ic,f}(t) = 1, \quad x_{ic,f}(0) = 1, \quad (4.16d)$$

$$\mathbf{x}_{iv,f}(t) = \mathbf{0}_{q_f}. \quad (4.16e)$$

The shape constraints for the extents of inlet and the extents of reaction introduced in Lemma 3 and Lemma 4 are valid for heterogeneous reaction systems. For the vessel extent of mass transfer, $\mathbf{x}_{m,f}(t)$, we have:

Lemma 8: (Extent of mass transfer)

The extent of mass transfer $x_{m,f,j}$ is (i) a nonnegative nondecreasing function if the j th species is added to phase F via mass transfer, (ii) a nonpositive nonincreasing function if the j th species is removed from phase F via mass transfer, and (iii) a concave (convex) function if the corresponding mass transfer rates is nonnegative nonincreasing (nondecreasing).

Proof: The proof follows from Lemmas 3 and 4. \square

Reactors with outlet

For reactors with an outlet flow, shape constraints for the extent of initial conditions are given in Lemma 5. Furthermore, the generalized shape constraints for the extents of reactions and inlets derived for homogeneous reaction systems are still valid. The corresponding constraints for extents of mass transfer can be written as:

Proposition 4.3 (Generalized shape constraints for the extents of mass transfer)

The generalized shape constraint for the species j in phase F that transfers to that phase ($\zeta_{f,j} \geq 0$) is:

$$\dot{x}_{m,f,j}(t) + \omega_f(t) x_{m,f,j}(t) \geq 0. \quad (4.17)$$

The generalized shape constraint for the species j in phase F that transfers to another phase ($\zeta_{f,j} \leq 0$) is:

$$\dot{x}_{m,f,j}(t) + \omega_f(t) x_{m,f,j}(t) \leq 0, \quad (4.18)$$

with $x_{m,f,j}$ the extent of mass transfer of the species j .

Proof: Eqs. (4.17) and (4.18) are obtained by re-arranging Eq. (3.39b) and using the fact that $\zeta_{f,j}(t) \geq 0$ for a species that transfers to phase F and $\zeta_{f,j}(t) \leq 0$ for a species that leaves phase F.

If measurements of the outlet flowrate and of the mass are available, the values $\omega_f(t_h) = \frac{u_{out,f}(t_h)}{m_f(t_h)}$, $\forall h = 1, 2, \dots, H$, are known, and Eqs. (4.17) and (4.18) can be evaluated by discretization of the derivative terms at the sampling instants:

$$\frac{x_{m,f,j}(t_h) - x_{m,f,j}(t_{h-1})}{t_h - t_{h-1}} + \omega_f(t_{h-1}) x_{m,f,j}(t_{h-1}) \geq 0, \quad \text{if } \zeta_{f,j} \geq 0, \quad (4.19a)$$

$$\frac{x_{m,f,j}(t_h) - x_{m,f,j}(t_{h-1})}{t_h - t_{h-1}} + \omega_f(t_{h-1}) x_{m,f,j}(t_{h-1}) \leq 0, \quad \text{if } \zeta_{f,j} \leq 0, \quad \forall h = 2, \dots, H. \quad (4.19b)$$

On the other hand, if measurements of the outlet flowrate and the mass are not available, $\omega_f(t_{h-1})$ in Eqs. (4.19a) and (4.19b) can be replaced by its expression in Eq. (4.10), which allows rewriting these shape constraints as:

$$x_{m,f,j}(t_h) - \frac{x_{ic,f}(t_h)}{x_{ic,f}(t_{h-1})} x_{m,f,j}(t_{h-1}) \geq 0, \quad \text{if } \zeta_{f,j} \geq 0, \quad (4.20a)$$

$$x_{m,f,j}(t_h) - \frac{x_{ic,f}(t_h)}{x_{ic,f}(t_{h-1})} x_{m,f,j}(t_{h-1}) \leq 0, \quad \text{if } \zeta_{f,j} \leq 0, \quad \forall h = 2, \dots, H. \quad (4.20b)$$

4.3 Data Reconciliation

Data reconciliation of reaction systems uses both redundancies among variables and shape constraints to improve the accuracy of measurements. These redundancies can be expressed as algebraic invariant relationships that are extracted from the balance equations. They can also be seen as process constraints among measured variables. The following assumptions are made regarding the chemical reaction system:

Assumption 4.1

The number of species S is greater than or equal to the dimensionality d .

Assumption 4.2

The numbers of moles of all S species are measured at all sampling times.

Assumption 4.3

\mathbf{N} , \mathbf{W}_{in} and \mathbf{n}_0 are perfectly known.

Assumption 4.1 ensures that all the vessel extents can be computed from the measured numbers of moles via the transformation matrix \mathcal{T} . Chapter 6 discusses the case of a subset of measured species. Assumption 4.3 ensures that the transformation matrix \mathcal{T} can be constructed without any uncertainty.

In this section, for the sake of simplicity, DR problems are formulated in terms of numbers of moles and extents for homogeneous reaction systems. The extension of the data reconciliation problem to the heterogeneous reaction systems is given in Appendix A.

4.3.1 Reconciliation using numbers of moles

Data reconciliation in terms of numbers of moles is formulated as a weighted least-squares optimization problem constrained by the invariant relationships \mathcal{I} , the knowledge-based constraints \mathcal{K}_n , the measurement-based constraints \mathcal{M}_n , and positivity constraints.

The reconciliation problem can be formulated as follows:

$$\hat{\mathbf{n}}(t_{1:H}) = \arg \min_{\mathbf{n}(t_{1:H})} \sum_{h=1}^H (\tilde{\mathbf{n}}(t_h) - \mathbf{n}(t_h))^T \mathbf{W}(t_h) (\tilde{\mathbf{n}}(t_h) - \mathbf{n}(t_h)) \quad (4.21)$$

$$\begin{aligned} \text{s.t.} \quad & \mathcal{I}(\mathbf{n}(t_{1:H})) = \mathbf{0}_{q \times H} && \text{(Invariant constraints)} \\ & \mathcal{K}_n(\mathbf{n}(t_{1:H})) \leq \mathbf{0}_{k_n \times H} && \text{(Knowledge-based constraints)} \\ & \mathcal{M}_n(\mathbf{n}(t_{1:H})) \leq \mathbf{0}_{m_n \times H} && \text{(Measurement-based constraints)} \\ & \mathbf{n}(t_{1:H}) \geq \mathbf{0}_{S \times H}, && \text{(Positivity constraints)} \end{aligned}$$

with $\hat{\mathbf{n}}(t_{1:H}) = [\hat{\mathbf{n}}(t_1), \dots, \hat{\mathbf{n}}(t_H)]$ the $S \times H$ dimensional reconciled numbers of moles, $\hat{\mathbf{n}}(t_h)$ the vector of reconciled numbers of moles at time t_h , $t_h \in \{t_1, \dots, t_H\}$, and $\mathbf{W}(t_h) = \Sigma_n^{-1}(t_h)$ the weighting matrix at time t_h .

Remark 4.2

When assumption 4.3 is not valid due to uncertain initial conditions, the matrix containing the invariant relationships, namely, \mathbf{P} , is uncertain. In such a case, the DR problem can be solved using the procedure described in [55].

4.3.2 Reconciliation using extents

Data reconciliation in terms of extents is formulated as a weighted least-squares optimization problem constrained by the knowledge-based \mathcal{K}_x , the measurement-based constraints \mathcal{M}_x , and positivity constraints. The reconciliation problem can be formulated as follows:

$$\hat{\mathbf{x}}(t_{1:H}) = \arg \min_{\mathbf{x}(t_{1:H})} \sum_{h=1}^H (\tilde{\mathbf{n}}(t_h) - \mathbf{n}(t_h))^T \mathbf{W}(t_h) (\tilde{\mathbf{n}}(t_h) - \mathbf{n}(t_h)) \quad (4.22)$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{n}(t_{1:H}) = \mathbf{B} \mathbf{x}(t_{1:H}) \\ & \mathcal{K}_x(\mathbf{x}(t_{1:H})) \leq \mathbf{0}_{k_x \times H} && \text{(Knowledge-based constraints)} \\ & \mathcal{M}_x(\mathbf{x}(t_{1:H})) \leq \mathbf{0}_{m_x \times H} && \text{(Measurement-based constraints)} \\ & \mathbf{n}(t_{1:H}) \geq \mathbf{0}_{S \times H} \text{ and } \mathbf{x}(t_{1:H}) \geq \mathbf{0}_{d \times H}, && \text{(Positivity constraints)} \end{aligned}$$

with $\hat{\mathbf{x}}(t_{1:H}) := \begin{bmatrix} \hat{\mathbf{x}}_r(t_{1:H}) \\ \hat{\mathbf{x}}_{in}(t_{1:H}) \\ \hat{\mathbf{x}}_{ic}(t_{1:H}) \end{bmatrix}$, and $\hat{\mathbf{x}}_r(t_{1:H}) = [\hat{\mathbf{x}}_r(t_1), \dots, \hat{\mathbf{x}}_r(t_H)]$, $\hat{\mathbf{x}}_{in}(t_{1:H}) = [\hat{\mathbf{x}}_{in}(t_1), \dots, \hat{\mathbf{x}}_{in}(t_H)]$ and $\hat{\mathbf{x}}_{ic}(t_{1:H}) = [\hat{\mathbf{x}}_{ic}(t_1), \dots, \hat{\mathbf{x}}_{ic}(t_H)]$ the sequences of reconciled extents of reaction, inlet and initial conditions. Note that the q invariant constraints are implicitly satisfied in the formulation of Eq. (4.22) since the invariants \mathbf{x}_{iv} are set to zero (and hence absent) in $\hat{\mathbf{x}}(t_{1:H})$.

4.3.3 Additional processing

4.3.3.1 Regularization

If desired, the solution delivered by DR formulated in terms of numbers of moles or extents can be smoothed by adding a penalty term to the objective function [60]. This can be achieved by modifying the objective functions of Eqs (4.21) and (4.22) as follows:

$$\sum_{h=1}^H (\tilde{\mathbf{n}}(t_h) - \mathbf{n}(t_h))^T \mathbf{W}(t_h) (\tilde{\mathbf{n}}(t_h) - \mathbf{n}(t_h)) + \sum_{h=1}^H J(t_h) \quad (4.23)$$

where $J(t_h) = \mathbf{1}_S^T \mathbf{\Lambda}_n \ddot{\mathbf{n}}(t_h)$, with $\mathbf{\Lambda}_n$ an $S \times S$ diagonal matrix of regularization parameters, is the penalty term used in Eq. (4.21), and $J(t_h) = \mathbf{1}_d^T \mathbf{\Lambda}_x \ddot{\mathbf{x}}(t_h)$, with $\mathbf{\Lambda}_x$ a $(d \times d)$ – diagonal matrix of regularization parameters, is the penalty term used in Eq. (4.22).

4.3.3.2 Variance of reconciled estimates

A number of different techniques exist for determining the variance of reconciled estimates [43]. Here, the variance of the reconciled estimates is approximated by considering only the equality constraints of the reconciliation problem. The variance-covariance matrix can also be approximated by sampling techniques [64]. One possible sampling technique consists in (i) computing the normal distribution centered around the measured numbers of moles or estimated extents at each time instant, and (ii) sampling the region of the normal distribution that satisfies the inequality constraints. Note that,

since the region of a normal distribution satisfying the inequality constraints no longer follows a Gaussian distribution, this sampling technique delivers the first and second moments of the reconciled estimates following unknown probability distributions [50].

4.4 Simulated Examples

In this section, the DR approaches introduced in Section 4.3 are illustrated in terms of both numbers of moles and extents for a homogeneous and a heterogeneous reaction system.

4.4.1 Homogeneous reaction systems

Consider the acetoacetylation of pyrrole system given in Example 2.3 and Section 3.3.1. This homogeneous reaction system will be investigated both in an open reactor with inlet and outlet streams and in a semi-batch reactor.

4.4.1.1 Open Reactor with Inlet and Outlet Streams

The inlet stream contains the species A, B and K, with the composition given by \mathbf{W}_{in} in Eq. (3.16) and a constant inlet flowrate of 2 g min^{-1} . The reactor has a constant outlet flowrate of 2 g min^{-1} . The reactor initially contains 2 mol of A, 5 mol of B and 0.5 mol of catalyst K, for an initial volume of 0.593 L. Assuming constant density, the volume remains constant at 0.593 L.

Measurements are taken every 0.5 min for 30 min and each measured number of moles is corrupted with additive and independent zero-mean Gaussian noise. The variance-covariance matrix Σ_n is constant and assumed to be:

$$\Sigma_n = \text{diag}([10^{-2}, 6 \times 10^{-2}, 2 \times 10^{-3}, 9 \times 10^{-3}, 10^{-4}, 8 \times 10^{-7}, 6 \times 10^{-4}]),$$

where $\text{diag}(\cdot)$ converts a vector into a diagonal matrix for the set of species $\mathcal{S} = \{ \text{A, B, C, D, E, F, K} \}$.

The flowrates and the initial conditions inside the reactor are assumed to be perfectly known, without bias or measurement noise.

Knowledge-based constraints

Data are first reconciled using only knowledge-based constraints.

In terms of numbers of moles

Since the reaction system is operated in an open reactor with constant volume and constant density, the dimensionality of the system is $d = R + p$. With $S = 7$, $R = 4$

and $p = 1$, one has $d = R + p = 5$, and there exist $q = S - d = 2$ invariants, whose mathematical form is described in Example 2.3.

This reaction system does not have non-added non-produced species. Moreover, due to the presence of the outlet flow, it does not have non-removed non-consumed species either. This means that, for this operating mode, it is not possible to impose any knowledge-based constraints in terms of numbers of moles. Hence, DR is performed with two invariant relationships and positivity constraints on all the numbers of moles. The residual sum of squares between the true values and the reconciled values are tabulated in Table 4.1 (columns labeled \mathcal{K}).

In terms of extents

In this case, since the flowrates are perfectly known, the extents of inlet and of initial conditions can be computed using Eqs (3.10b) and (3.10c). Due to the outlet term, generalized shape constraints are used for the extents of reaction as per Eq. (4.11a). The residual sum of squares between the true numbers of moles and the numbers of moles reconciled in terms of extents are tabulated in Table 4.1 (columns labeled \mathcal{K}). Fig. 4.1 shows the numbers of moles reconciled in terms of numbers of moles and extents using only knowledge-based constraints.

Table 4.1 Sum of squared errors for the simulated numbers of moles (referred to as measured quantities, $\tilde{\mathbf{n}}$) and the numbers of moles reconciled in terms of the numbers of moles \mathbf{n} and the extents \mathbf{x} using only knowledge-based constraints \mathcal{K} , and using both knowledge-based and measurement-based constraints $\mathcal{K} + \mathcal{M}$, for the pyrrole system in an open reactor with inlet and outlet streams.

Species	Measured $\tilde{\mathbf{n}}$	Reconciled (\mathcal{K})		Reconciled ($\mathcal{K} + \mathcal{M}$)	
		via \mathbf{n}	via \mathbf{x}	via \mathbf{n}	via \mathbf{x}
A	0.331	0.340	0.019	0.189	0.007
B	4.045	1.199	0.157	0.581	0.042
C	0.145	0.116	0.019	0.040	0.007
D	0.414	0.235	0.036	0.164	0.016
E	0.009	0.009	0.002	0.001	0.001
F	0.000	0.000	0.000	0.000	0.000
K	0.027	0.000	0.000	0.000	0.000

Measurement-based constraints

Data are then reconciled using both knowledge-based and measurement-based constraints.

In terms of numbers of moles

Monotonicity constraints as well as curvature shape constraints are estimated using the reconciled numbers of moles via the procedure outlined in Section 4.1.1.2. The identified

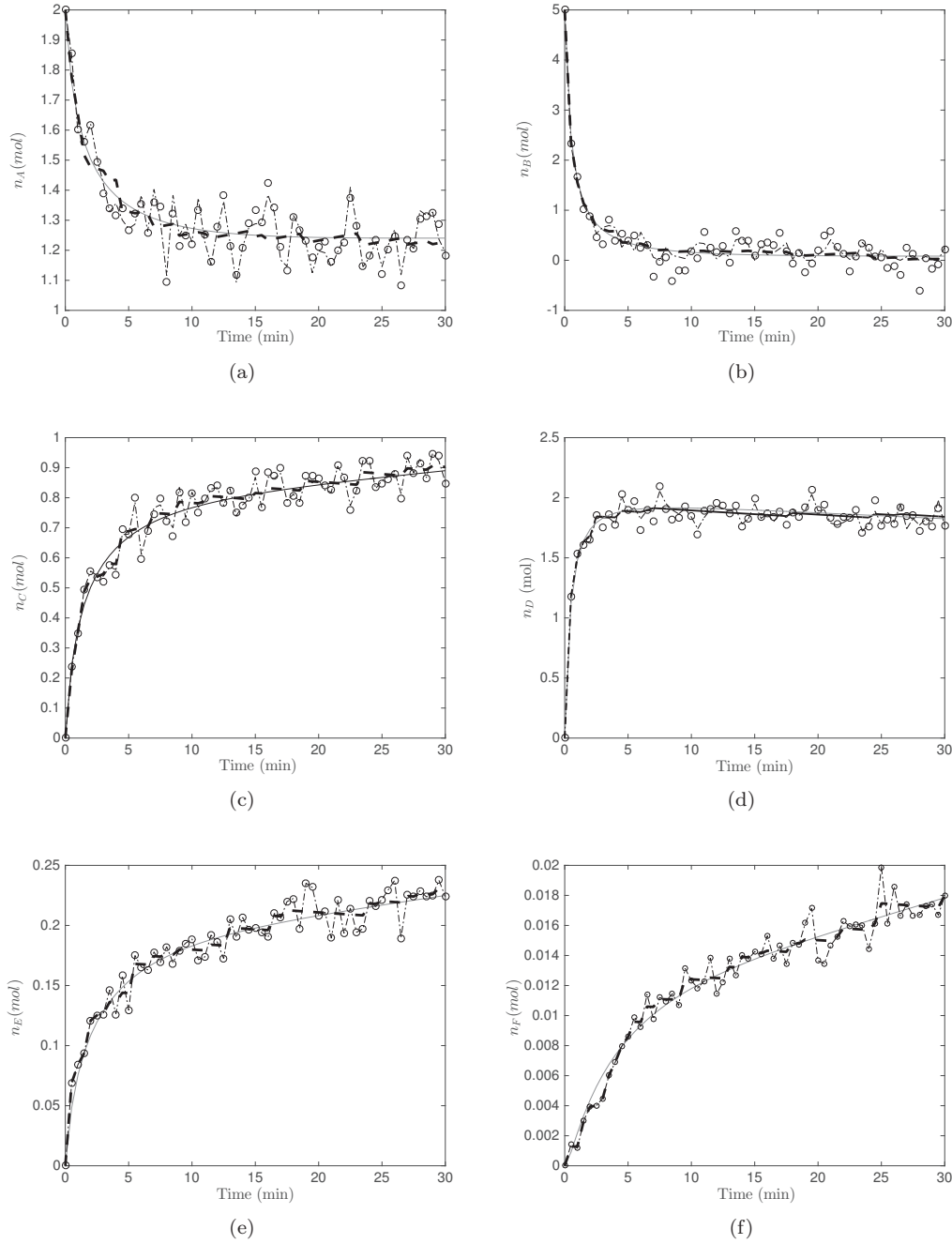


Figure 4.1 True (—) and measured (o) numbers of moles compared to the numbers of moles reconciled via \mathbf{n} (---) and \mathbf{x} (---) using only the knowledge-based constraints for the pyrrole system in an open reactor with inlet and outlet streams.

constraints are listed in Table 4.2. The residual sum of squares with these additional shape constraints are shown in Table 4.1 (columns labeled $\mathcal{K} + \mathcal{M}$).

Table 4.2 Measurement-based constraints (\mathcal{M}) identified using the procedure described in Section 4.1.1.2, for the pyrrole system in an open reactor with inlet and outlet streams. The symbols + and – indicate monotonically increasing and decreasing behaviors, whereas symbols \smile and \frown indicate convex and concave behaviors, respectively.

Species	Monotonicity	Until	Shape	Until
A	–	15.0 min	\smile	15.0 min
B	–	16.5 min	\smile	14.5 min
C	+	22.0 min	\frown	22.0 min
D	+	13.5 min	\frown	13.5 min
E	+	23.0 min	\frown	30.0 min
F	+	27.0 min	\frown	30.0 min
K	none	–	none	–
$x_{r,1}$	+(a)	30.0 min	\frown	30.0 min
$x_{r,2}$	+(a)	30.0 min	\frown	30.0 min
$x_{r,3}$	+(a)	30.0 min	\frown	30.0 min
$x_{r,4}$	+(a)	30.0 min	\frown	30.0 min

(a): the extents of reaction follow the generalized shape constraint in Eq. (4.8)

In terms of extents

Concave and convex shape constraints on the extents of reactions are estimated via the procedure of Section 4.1.1.2 from extents reconciled using only the knowledge-based constraints. In this case, all four extents of reaction are identified to be concave for all times. The residual sum of squares with these additional shape constraints are shown in Table 4.1 (columns labeled $\mathcal{K} + \mathcal{M}$).

It can be seen from the residual sum of squares given in Table 4.1 that DR in terms of extents leads to significantly better reconciliation than the formulation in terms of numbers of moles. This is due to the additional knowledge-based constraints that can be imposed in the extent formulation compared to the absence of such constraints in the number-of-moles formulation. It is also seen that the addition of the measurement-based curvature constraints further improves the reconciliation in both formulations. Fig. 4.2 shows the numbers of moles reconciled in terms of numbers of moles and extents using both knowledge-based and measurement-based constraints.

4.4.1.2 Semi-Batch Reactor

Consider the pyrrole system in a semi-batch reactor with an inlet stream containing pure B ($w_{in,B} = \frac{1}{M_{w,B}}$ with $M_{w,B} = 84 \text{ g mol}^{-1}$) with a flowrate of 5 g min^{-1} . The reactor

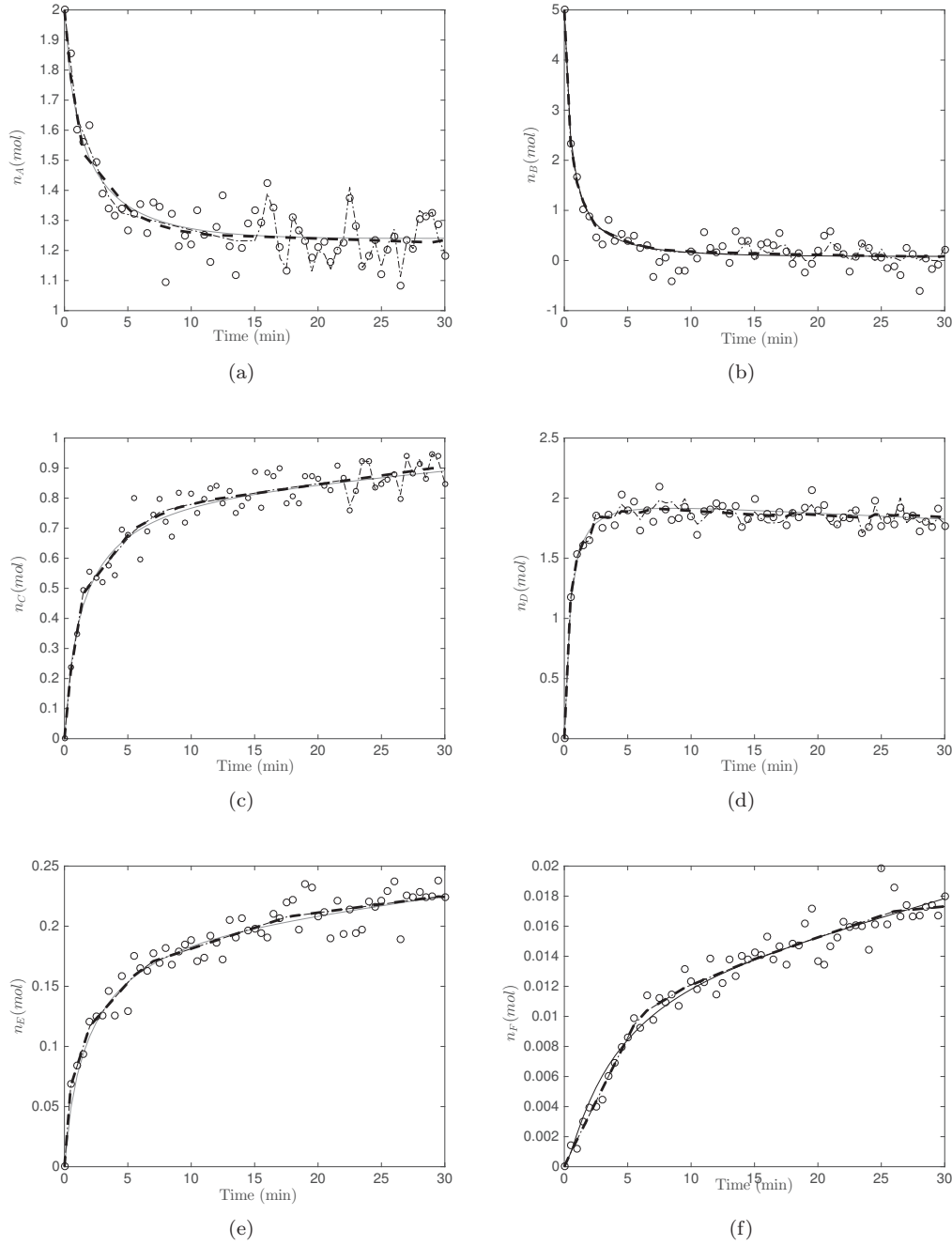


Figure 4.2 True (—) and measured (o) numbers of moles compared to the numbers of moles reconciled via \mathbf{n} (— · —) and \mathbf{x} (— —) using the knowledge-based and measurement-based constraints for the pyrrole system in an open reactor with inlet and outlet streams.

is initially loaded with the same amounts of material as the open reactor discussed in Section 4.4.1.1.

In this case, only the subset $\mathcal{S}_a = \{A, B, C, D, E\}$ is assumed to be measured, and their numbers of moles are corrupted with additive and independent zero-mean Gaussian noise. The variance-covariance matrix Σ_{n_a} is constant and assumed to be:

$$\Sigma_{n_a} = \text{diag}([10^{-2}, 6 \times 10^{-2}, 3.4 \times 10^{-3}, 1.4 \times 10^{-2}, 4 \times 10^{-4}]) .$$

Measurements of the \mathcal{S}_a species are taken every 0.5 min for 30 min. The inlet flowrate is assumed to be unknown in this case.

Knowledge-based constraints

Data are first reconciled using only knowledge-based constraints.

In terms of numbers of moles

Since the reaction system takes place in a semi-batch reactor with $R = 4$ reactions and $p = 1$ inlet, the dimensionality of the system is $d = R + p = 5$. With $S = 7$ and $S_a = 5$, the number of invariants is $q = q_u = 2$ and $q_a = 0$, with the 2 invariant relationships being:

$$n_A(t) + n_C(t) + n_F(t) = n_{A0} + n_{C0} + n_{F0} \quad (4.24a)$$

$$n_K(t) = n_{K0}. \quad (4.24b)$$

The unmeasured species $\mathcal{S}_u = \{F, K\}$ are reconstructed using these invariant relationships and thus no invariant is available for DR. Under these experimental conditions, the set of non-added non-produced species that are nonincreasing is $\mathcal{S}_{np} = \{A\}$, while the set of non-removed non-consumed species that are nondecreasing is $\mathcal{S}_{nc} = \{D, E, F\}$. Hence, DR is performed with three knowledge-based constraints for the measured species A, D, and E, one knowledge-based constraint for the unmeasured species F, and positivity of all the numbers of moles. The residual sum of squares between the true (noisy) and reconciled numbers of moles are tabulated in Table 4.3 (columns labeled \mathcal{K}).

In terms of extents

In terms of extents, due to the absence of an outlet stream, all 5 extents (4 extents of reaction and 1 extent of inlet) are nondecreasing. The residual sum of squares between the true (noisy) numbers of moles and the numbers of moles reconciled in terms of extents are tabulated in Table 4.3 (columns labeled \mathcal{K}).

Table 4.3 Sum of squared errors for the (noisy) simulated numbers of moles (referred to as measured quantities, $\tilde{\mathbf{n}}$) and the numbers of moles reconciled in terms of the numbers of moles \mathbf{n} and the extents \mathbf{x} using only knowledge-based constraints \mathcal{K} , and using both knowledge-based and measurement-based constraints $\mathcal{K} + \mathcal{M}$, for the pyrrole system in a semi-batch reactor.

Species	Measured $\tilde{\mathbf{n}}$	Reconciled (\mathcal{K})		Reconciled ($\mathcal{K} + \mathcal{M}$)	
		via \mathbf{n}	via \mathbf{x}	via \mathbf{n}	via \mathbf{x}
A	0.760	0.059	0.044	0.048	0.034
B	3.710	3.630	0.695	3.054	0.689
C	0.180	0.057	0.051	0.045	0.041
D	0.820	0.223	0.103	0.223	0.099
E	0.019	0.009	0.008	0.006	0.005

Measurement-based constraints

Data are then reconciled using both knowledge-based constraints and measurement-based constraints.

In terms of numbers of moles

Monotonicity constraints on the species B and C, as well as curvature shape constraints on all the other measured species are identified using the reconciled numbers of moles via the procedure outlined in Section 4.1.1.2. Note that the monotonic behavior of the species A, D and E is known *a priori* from the knowledge-based constraints, as $A \in \mathcal{S}_{np}$, and both D and E $\in \mathcal{S}_{nc}$. The identified constraints are listed in Table 4.4. The residual sum of squares with these additional shape constraints are shown in Table 4.3 (columns labeled $\mathcal{K} + \mathcal{M}$).

In terms of extents

Concave and convex shape constraints on the extents of reactions were estimated via the procedure of Section 4.1.1.2 from the extents reconciled using only the knowledge-based constraints as shown in Table 4.4. The residual sum of squares with these additional shape constraints are shown in Table 4.3 (columns labeled $\mathcal{K} + \mathcal{M}$).

Similarly to the case of open reactors, DR in semi-batch reactors is more efficient in terms of extents than in terms of numbers of moles. In this example, this is due to the presence of knowledge-based constraints on all five extents compared to *a priori* constraints on just three species. Also, the measurement-based constraints further improves DR in both formulations.

Table 4.4 Measurement-based constraints (\mathcal{M}) identified using the reconciled numbers of moles and reconciled extents using the procedure described in Section 4.1.1.2, for the pyrrole system in a semi-batch reactor. The symbols + and – indicate monotonically increasing and decreasing behaviors, whereas symbols \smile and \frown indicate convex and concave behaviors, respectively.

Species	Monotonicity	Until	Shape	Until
A	– (a)	30.0 min	\smile	10.5 min
B	–	7.5 min	\smile	4.0 min
C	+	23.0 min	\frown	6.0 min
D	+	30.0 min	none	–
E	+	30.0 min	\frown	14.0 min
$x_{r,1}$	+	30.0 min	\frown	14.0 min
$x_{r,2}$	+	30.0 min	none	–
$x_{r,3}$	+	30.0 min	\frown	14.0 min
$x_{r,4}$	+	30.0 min	none	–
x_{in}	+	30.0 min	none	–

(a): known *a priori* from knowledge-based constraints

4.4.2 Heterogeneous reaction systems

Consider the chlorination of butanoic acid (BA) system introduced in Example 2.4 and Section 3.3.2. For the simulation, 10 kmol of butanoic acid, a small amount of MBA and 100 kmol of solvent are initially loaded in the reactor. The reactor is operated in semi-batch mode and continuously fed with chlorine gas with the mass flowrate $u_{in,g} = 972$ kg h^{–1}. The reactor volume V_r is constant at 9 m³, with the gas and liquid phases initially occupying 1.844 m³ and 7.156 m³, respectively. Numbers of moles of all the species in both phases are measured every minute for 50 min. The measured numbers of moles are corrupted with zero mean Gaussian white noise with constant variance-covariance matrices:

$$\Sigma_{n,l} = \text{diag}([0.4225, 0.0005, 0.4224, 0.2125, 0.0000]),$$

$$\Sigma_{n,g} = \text{diag}([0.0002, 0.0363])$$

The flowrates and the initial conditions inside the reactor are assumed to be perfectly known, without bias or measurement noise. For this case study, we illustrate the data reconciliation procedure using only the knowledge based constraints.

Knowledge-based constraints

In terms of numbers of moles

The liquid phase of the gas-liquid reaction system has $S = 5$ and dimensionality $d = R + p = 4$. This phase has a single invariant relationship, whose mathematical form can be described as:

$$n_{BA,l}(t) + n_{MBA,l}(t) + n_{DBA,l}(t) = n_{BA,l0} + n_{MBA,l0} + n_{DBA,l0}$$

In this phase, species BA belongs to the set of non-added non-produced species (monotonically decreasing), while species MBA and DBA belong to the set of non-removed non-consumed species (monotonically increasing). This means that, for this operating mode, knowledge-based constraints can be imposed on three of the five species. For the gas phase G with $S = 2$ species and $d = 3$, there exists no invariant relationships. However the species HCl in the gas phase belongs to the set of non-removed non-consumed species (monotonically increasing). The DR problem is solved for each phase and the residual sum of squares between the true values and the reconciled values are tabulated in Table 4.5.

In terms of extents

In phase L, due to the absence of the outlet stream, both extents of reaction are monotonically decreasing functions. Also, from Lemma 8, it follows that the extent of mass transfer corresponding to species Cl_2 is monotonically increasing (since it is added), while the extent of mass of mass transfer corresponding to species HCl is monotonically decreasing (since it is removed). For phase G, the extent of inlet is computed using Eq. (3.39c). The constraints on the extents of mass transfer in phase G is opposite to the constraints in phase L, i.e. extent of mass transfer due to HCl is increasing while extent corresponding to Cl_2 is decreasing. The residual sum of squares between the true numbers of moles and the numbers of moles reconciled in terms of extents are tabulated in Table 4.5. Fig. 4.3 and 4.4 shows the numbers of moles reconciled in terms of numbers of moles and extents using knowledge-based constraints for phase G and phase L respectively.

Table 4.5 shows that the data reconciliation in terms of vessel extents is more efficient than the data reconciliation in terms of numbers of moles. Similarly to the homogeneous case study, measurement-based constraints can be identified to further improve the accuracy of the reconciled estimates. Here the measurement-based constraints are not illustrated.

Table 4.5 Sum of squared errors for the simulated numbers of moles (referred to as measured quantities, $\tilde{\mathbf{n}}_f$) and the numbers of moles reconciled in terms of the numbers of moles \mathbf{n} and the extents \mathbf{x} using only knowledge-based constraints \mathcal{K} , for the chlorination of butanoic acid system in a semi-batch reactor with inlet stream in the gas phase.

Species	Measured $\tilde{\mathbf{n}}_f$	Reconciled (\mathcal{K})	
		via \mathbf{n}_f	via \mathbf{x}_f
$\text{Cl}_{2,g}$	0.0088	0.0088	0.0085
HCl_g	1.8622	0.5268	0.5240
BA_l	30.9739	8.6518	3.7265
$\text{Cl}_{2,l}$	0.0239	0.0160	0.0157
MBA_l	23.2403	8.4607	3.7267
HCl_l	14.3601	14.1282	3.8801
DBA_l	0.0000	0.0000	0.000

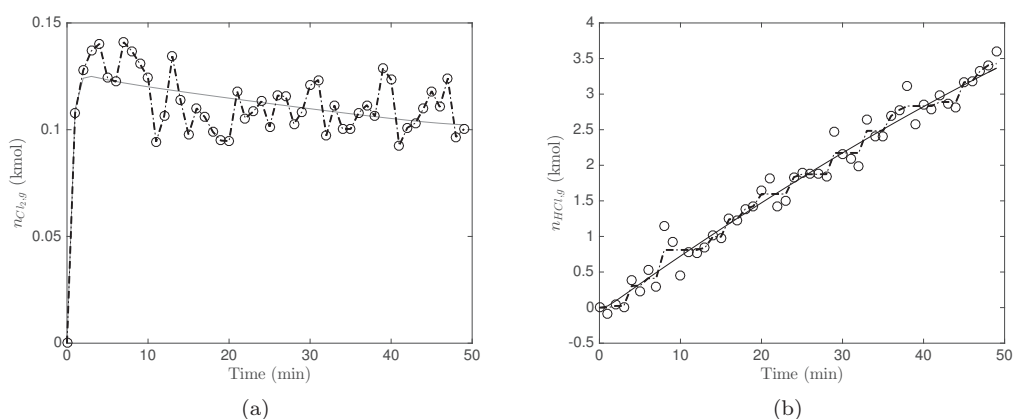


Figure 4.3 True (—) and measured (o) numbers of moles compared to the numbers of moles reconciled via \mathbf{n}_g (---) and \mathbf{x}_g (-.-) in phase G, using the knowledge-based constraints for the chlorination of butanoic acid system.

4.5 Summary

Data reconciliation uses redundancies expressed as relationships between state variables to reduce the noise in measured data. For chemical reaction systems, the invariant relationships can be used as constraints for data reconciliation. These relationships are algebraic constraints, since they do not contain information regarding the states about past and future time instants.

If a kinetic model is available, DR can be performed using well-established techniques, such as RNDDR or MHE. In the absence of a kinetic model, this chapter has shown how to identify *shape constraints* such as monotonicity and curvature (convexity/concavity) behavior on the numbers of moles, which are constraints relating the measurements over several time instants. These shape constraints can be used for DR. Conditions under

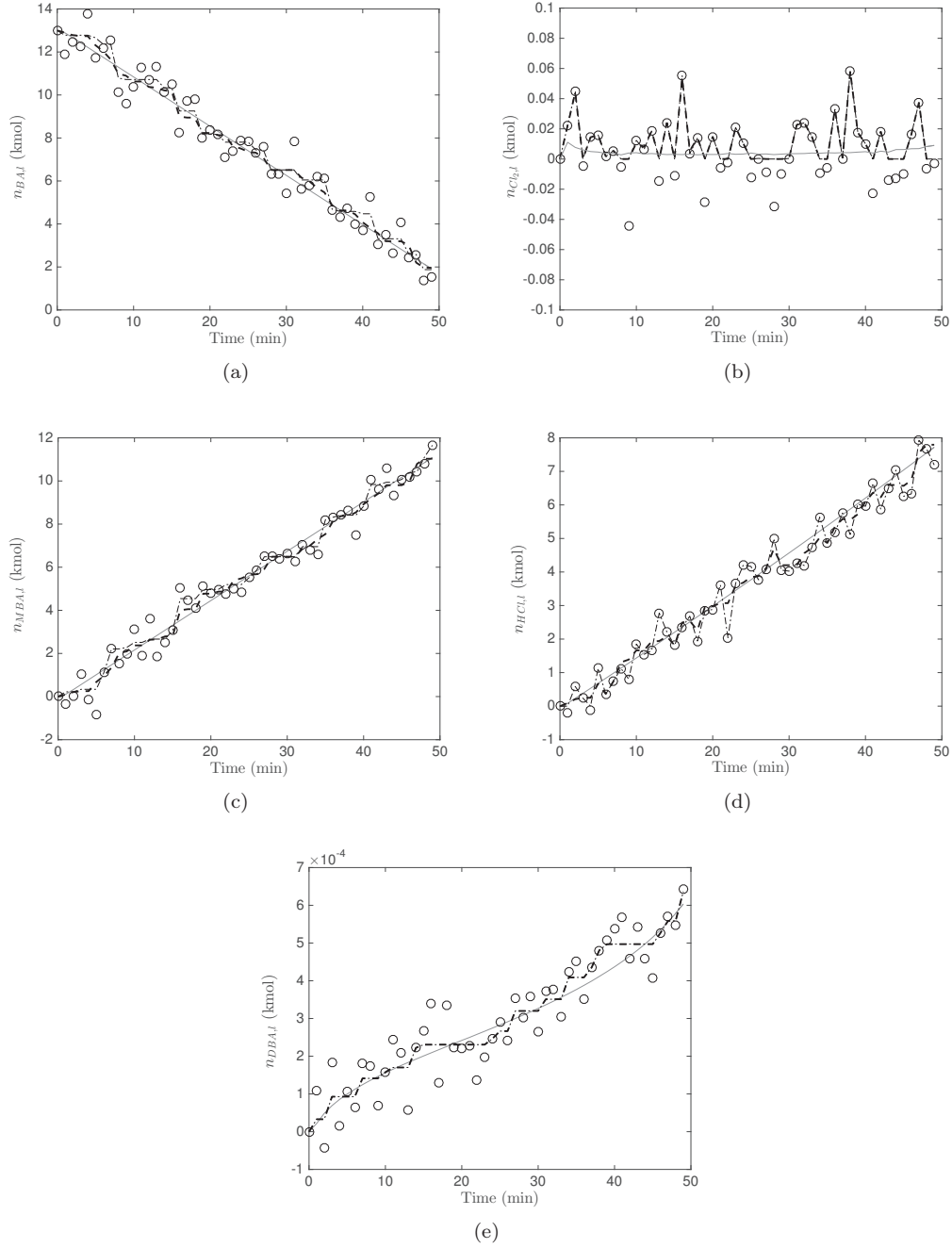


Figure 4.4 True (—) and measured (o) numbers of moles compared to the numbers of moles reconciled via \mathbf{n}_I (—) and \mathbf{x}_I (—) in phase L, using the knowledge-based constraints for the chlorination of butanoic acid system.

which these constraints are valid based on prior knowledge have been proposed. Furthermore, it was shown that an alternative formulation of the reaction system in terms of extents helps identify additional shape constraints, leading to better reconciled estimates. In addition, a procedure for deriving *shape constraints* based on measurements has also been developed.

Chapter 5

Sequential Model Identification

This chapter introduces a hybrid – sequential – approach that progressively identifies the kinetic model from a purely incremental to a purely simultaneous approach. In this sequential approach, the extents of all rate processes are ranked based on a pre-defined criterion, such as their variance (highest to lowest) or their signal-to-noise ratio (lowest to highest). The method starts with the identification (model structure and parameter values) of the *first-ranked* rate process using measured extents for all the other rates. Note that this step is a purely extent-based incremental step. The model structure of the first rate process is then fixed for all other subsequent identification problems. As the next step, the model structure of the *second-ranked* rate process is identified, and the parameters of the *first and second* rate models are simultaneously adjusted using measured extents for the remaining rates. This procedure is repeated for all rate processes by fixing the model structures of all the previously identified rates and using measured extents for the remaining rate processes. The identification of the last rate process is of the purely simultaneous type since all rate parameters are estimated simultaneously using the model structures that have been determined previously and the model candidates for the last rate. Assuming M model candidates to be tested per rate process, this hybrid approach requires $M \cdot N$ regressions to identify the models of all N rate processes.

This chapter is organized as follows. Section 5.1.2, briefly reviews the the extent-based incremental model identification (IMI_n) introduced by Bonvin and co-workers [4, 15]. This section also introduces another variant of the extent-based incremental approach, IMI_x , where the model identification is carried out entirely in terms of vessel extents. Section 5.1.3 introduces the sequential model identification procedure for a homogeneous reaction system. Section 5.1.4 discusses the structural and interpolation biases in the incremental and sequential approaches. Finally, the sequential approach is illustrated via the simulation of a homogeneous reaction system in Section 5.2, and Section 5.3 concludes the chapter. For pedagogical reasons, this chapter discusses only the homogeneous reaction system, the sequential model identification procedure for the heterogeneous reaction system is provided in Appendix B.

5.1 Homogeneous Reaction Systems

This section details the procedure for identifying the models for all the R reactions in a homogeneous reaction system using the extent-based incremental and sequential approaches.

5.1.1 Model identification

Assumption 5.1

Measurements of all S species are available at all sampling times $t_h, \forall h = 1, \dots, H$.

Assumption 5.2

The initial conditions \mathbf{n}_0 at $t_1 = 0$ are perfectly known.

Assumption 5.3

If the reactor has inlet and outlet streams then their corresponding flowrates \mathbf{u}_{in} and u_{out} are perfectly known.

The measured numbers of moles at other time instants are corrupted by independent Gaussian noise of zero mean and known variance-covariance matrix $\Sigma_n(t_h)$, such that:

$$\tilde{\mathbf{n}}(t_h) = \mathcal{N}(\mathbf{n}_{true}(t_h), \Sigma_n(t_h)), \quad \forall h = 2, \dots, H,$$

with \mathbf{n}_{true} and $\tilde{\mathbf{n}}$ the true and measured values of the numbers of moles. The quality of the measured numbers of moles are improved via the data reconciliation procedure introduced in Chapter 4:

$$\hat{\mathbf{n}}(t_h) = \mathcal{N}(\mathbf{n}_{true}(t_h), \Sigma_{\hat{\mathbf{n}}}(t_h)), \quad \forall h = 2, \dots, H,$$

where $\hat{\mathbf{n}}$ denotes the reconciled numbers of moles. For the purpose of simplifying the notation, let us define $\boldsymbol{\mu} := \begin{bmatrix} \mathbf{x}_{in} \\ x_{ic} \end{bmatrix}$, which allows writing the dynamics of the extents of inlet and initial conditions as:

$$\dot{\boldsymbol{\mu}}(t) = \begin{bmatrix} \mathbf{u}_{in} \\ 0 \end{bmatrix} - \omega(t) \boldsymbol{\mu}(t), \quad \boldsymbol{\mu}(0) = \begin{bmatrix} \mathbf{0}_p \\ 1 \end{bmatrix}. \quad (5.1)$$

For the extents of reaction, $\mathbf{x}_r(t)$, the reaction rates $\mathbf{r}_v(t)$ are treated as time signals that could be modeled as unknown functions of the numbers of moles or vessel extents, namely, $\mathbf{r}_v(t) := V(t)\varphi_n(\mathbf{n}(t), \boldsymbol{\theta})$ or $\mathbf{r}_v(t) := V(t)\varphi_x(\mathbf{x}_r(t), \boldsymbol{\mu}(t), \boldsymbol{\theta})$. The extents of reaction can then be represented as:

$$\dot{\mathbf{x}}_r(t) = V(t) \varphi_n(\mathbf{n}(t), \boldsymbol{\theta}) - \omega(t) \mathbf{x}_r(t) \quad (5.2)$$

$$= V(t) \varphi_x(\mathbf{x}_r(t), \boldsymbol{\mu}(t), \boldsymbol{\theta}) - \omega(t) \mathbf{x}_r(t), \quad \mathbf{x}_r(0) = \mathbf{0}_R \quad (5.3)$$

Also, let us introduce $\mathcal{M}_n = \{\mathcal{M}_{n,1}, \dots, \mathcal{M}_{n,R}\}$ the set of all $M = \sum_{k=1}^R M_k$ candidate models for the R reactions expressed in terms of numbers of moles, and $\mathcal{M}_x = \{\mathcal{M}_{x,1}, \dots, \mathcal{M}_{x,R}\}$ the corresponding set in terms of extents. In particular, $\mathcal{M}_{n,k} = \{\varphi_{n,k}^{(1)}, \varphi_{n,k}^{(2)}, \dots, \varphi_{n,k}^{(M_k)}\}$ and $\mathcal{M}_{x,k} = \{\varphi_{x,k}^{(1)}, \varphi_{x,k}^{(2)}, \dots, \varphi_{x,k}^{(M_k)}\}$ denote the sets of M_k candidate models for the k th reaction in terms of numbers of moles and extents.

5.1.2 Extent-based Incremental Model Identification (IMI)

This extent-based incremental approach, which has been introduced by Bonvin and coworkers [4, 15] is a three-step process.

Rate expressions expressed in terms of numbers of moles

The procedure for the extent-based incremental approach in terms of numbers of moles is detailed in Algorithm 1. The integration of the candidate models in Step 3 involves the knowledge of the reconciled numbers of moles at all times, which is obtained by interpolation of the numbers of moles $\hat{\mathbf{n}}$ available at the discrete time instants $\{t_1, t_2, \dots, t_H\}$. The initial guesses for the simultaneous estimation of $\boldsymbol{\theta}$ in Eq. (5.5) are usually taken as the estimated model parameters obtained via Eq. (5.4). For the purpose of comparison, this variant of the incremental model identification with rate expressions expressed in terms of numbers of moles is labelled IMI_n.

Rate expressions expressed in terms of extents

The procedure labeled IMI_x is detailed in Algorithm 2. The rate function $\varphi_{n,k}^{(m_k)}(\hat{\mathbf{n}}(t), \boldsymbol{\theta}_k^{(m_k)})$ in Eq. (5.4) can also be formulated in terms of the extents as $\varphi_{x,k}^{(m_k)}(x_{r,k}(t), \hat{\mathbf{x}}_{r,\mathcal{J}}(t), \boldsymbol{\mu}(t), \boldsymbol{\theta}_k^{(m_k)})$, as shown in Eq. (5.6). In such a case, $(R - 1)$ reconciled extents $\hat{\mathbf{x}}_{r,\mathcal{J}}$ (and not the reconciled numbers of moles) need to be interpolated, and the extent $x_{r,k}$ that is being modeled can be integrated without interpolation, just from the knowledge of its initial condition. The initial guesses for the simultaneous estimation of $\boldsymbol{\theta}$ in Eq. (5.7) are taken as the estimated model parameters obtained via Eq. (5.6).

5.1.3 Extent-based Sequential Model Identification (SMI)

The extent-based sequential model identification is also a multi-step approach. However, while the incremental approach models each reaction individually, the sequential approach uses the models identified in the previous steps to facilitate the identification of the next reactions, as detailed in Algorithm 3. The order selection in Step 3 can be based on several criteria. Here, two criteria are suggested:

Algorithm 1 Extent-based incremental model identification (variant IMI_n)

Step 1: Compute $\mathbf{x}_{in}(t)$, $\mathbf{x}_{ic}(t)$ and $\hat{\mathbf{n}}^{vRV}(t)$ using Eqs. (3.10b), (3.10c) and (3.29)

Step 2: Compute $\hat{\mathbf{x}}_r(t)$ for all R reactions using Eq. (3.32)

Step 3: The model identification task is carried out by postulating rate expressions for each reaction individually. The model identification problem can be solved in parallel for the R reactions as follows:

for $k = 1, \dots, R$

for $m_k = 1, \dots, M_k$

$$\boldsymbol{\theta}_k^{(m_k)*} = \arg \min_{\boldsymbol{\theta}_k^{(m_k)}} J(\boldsymbol{\theta}_k^{(m_k)})$$

$$\begin{aligned} \text{s.t.} \quad J &= \sum_{h=1}^H \left(\hat{x}_{r,k}(t_h) - x_{r,k}^{(m_k)}(t_h) \right) W_k(t_h) \left(\hat{x}_{r,k}(t_h) - x_{r,k}^{(m_k)}(t_h) \right) \\ \dot{x}_{r,k}^{(m_k)}(t) &= V(t) \varphi_{n,k}^{(m_k)}(\hat{\mathbf{n}}(t), \boldsymbol{\theta}_k^{(m_k)}) - \omega(t) x_{r,k}^{(m_k)}(t), \quad x_{r,k}^{(m_k)}(0) = 0. \end{aligned} \quad (5.4)$$

end for

$$m_k^* = \arg \min_{m_k} \left\{ J(\boldsymbol{\theta}_k^{(m_k)*}) \right\}_{m_k=1}^{M_k}$$

end for

with $W_k(t_h)$ the inverse of the variance $\sigma_{x_{r,k}}^2(t_h)$ of the k th measured reaction extent at time t_h . The model candidate with the least objective function $J(\boldsymbol{\theta}_k^{(m_k)*})$ is chosen as the best model $\varphi_{n,k}^{(m_k^*)}$ for that reaction along with the values of its corresponding parameters $\boldsymbol{\theta}_k^{(m_k^*)}$.

Step 4: Once the models of all reactions have been identified incrementally, a simultaneous parameter estimation problem is solved to obtain the parameter values in the maximum-likelihood sense:

$$\begin{aligned} \boldsymbol{\theta}^* &= \arg \min_{\boldsymbol{\theta}} \sum_{h=1}^H \left(\hat{\mathbf{x}}_r(t_h) - \mathbf{x}_r(t_h) \right)^T \mathbf{W}(t_h) \left(\hat{\mathbf{x}}_r(t_h) - \mathbf{x}_r(t_h) \right) \\ \text{s.t.} \quad \dot{\mathbf{x}}_r(t) &= V(t) \boldsymbol{\varphi}_n^{(m^*)}(\mathbf{n}(t), \boldsymbol{\theta}) - \omega(t) \mathbf{x}_r(t), \quad \mathbf{x}_r(0) = \mathbf{0}_R, \end{aligned} \quad (5.5)$$

with $\mathbf{W}(t_h) = \boldsymbol{\Sigma}_x^{-1}(t_h)$ the inverse of the variance-covariance matrix at time t_h , $\boldsymbol{\varphi}_n^{(m^*)} = [\varphi_{n,1}^{(m_1^*)}, \varphi_{n,2}^{(m_2^*)}, \dots, \varphi_{n,R}^{(m_R^*)}]^T$ the R rate models identified via the parallel incremental identification of Step 3, and $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_R^T]^T$ the adjustable model parameters of the identified rate models.

Algorithm 2 Extent-based incremental model identification (variant IMI_x)

Step 1: Compute $\mathbf{x}_{in}(t)$, $\mathbf{x}_{ic}(t)$ and $\hat{\mathbf{n}}^{vRV}(t)$ using Eqs. (3.10b), (3.10c) and (3.29)

Step 2: Compute $\hat{\mathbf{x}}_r(t)$ for all R reactions using Eq. (3.32)

Step 3: The model identification task is carried out by postulating rate expressions for each reaction individually. Let the subscript k denote the reaction being modeled, and \mathcal{J} the set of reactions that are not modeled. The model identification problem can be solved in parallel for the R reactions as follows:

for $k = 1, \dots, R$

for $m_k = 1, \dots, M_k$

$$\boldsymbol{\theta}_k^{(m_k)*} = \arg \min_{\boldsymbol{\theta}_k^{(m_k)}} J(\boldsymbol{\theta}_k^{(m_k)})$$

$$\begin{aligned} \text{s.t.} \quad J &= \sum_{h=1}^H \left(\hat{x}_{r,k}(t_h) - x_{r,k}^{(m_k)}(t_h) \right) W_k(t_h) \left(\hat{x}_{r,k}(t_h) - x_{r,k}^{(m_k)}(t_h) \right) \\ \dot{x}_{r,k}^{(m_k)}(t) &= V(t) \varphi_{x,k}^{(m_k)}(x_{r,k}(t), \hat{\mathbf{x}}_{r,\mathcal{J}}(t), \boldsymbol{\mu}(t), \boldsymbol{\theta}_k^{(m_k)}) - \omega(t) x_{r,k}^{(m_k)}(t), \\ x_{r,k}^{(m_k)}(0) &= 0. \end{aligned} \quad (5.6)$$

end for

$$m_k^* = \arg \min_{m_k} \left\{ J(\boldsymbol{\theta}_k^{(m_k)*}) \right\}_{m_k=1}^{M_k}$$

end for

The model candidate with the least objective function is chosen as the best model $\varphi_{n,k}^{(m_k^*)}$ for that reaction along with the values of its corresponding parameters $\boldsymbol{\theta}_k^{(m_k^*)*}$.

Step 4: Once the models of all reactions have been identified incrementally, a simultaneous parameter estimation problem is solved to obtain the parameter values in the maximum-likelihood sense:

$$\begin{aligned} \boldsymbol{\theta}^* &= \arg \min_{\boldsymbol{\theta}} \sum_{h=1}^H \left(\hat{\mathbf{x}}_r(t_h) - \mathbf{x}_r(t_h) \right)^T \mathbf{W}(t_h) \left(\hat{\mathbf{x}}_r(t_h) - \mathbf{x}_r(t_h) \right) \\ \text{s.t.} \quad \dot{\mathbf{x}}_r(t) &= V(t) \varphi_x^{(m^*)}(\mathbf{x}_r(t), \boldsymbol{\mu}(t), \boldsymbol{\theta}) - \omega(t) \mathbf{x}_r(t), \quad \mathbf{x}_r(0) = \mathbf{0}_R. \end{aligned} \quad (5.7)$$

- **Signal-to-Noise Ratio:** Compute the signal-to-noise ratio (SNR) for the extents to be modeled and order the extents in an increasing SNR, that is, the extent with the lowest SNR will be modeled first.
- **Noise power:** Rank the extents in decreasing order with respect to the noise variance $\boldsymbol{\Sigma}_x(t_h) = \mathcal{T} \boldsymbol{\Sigma}_{\hat{n}}(t_h) \mathcal{T}^T$, that is, the extent with the highest noise power will be modeled first.

Since the quality of model identification depends on the quality of the interpolation of noisy measurements, the interpolation error is reduced at each step of the method if the least noisy extents are interpolated first, or, equivalently, if the most noisy extents (low SNR or high noise power) are modeled first. The model identification problem in the sequential approach is formulated in terms of extents. The function $\varphi_{x,k}^{(m_k)}(x_{r,k}^{(m_k)}(t), \mathbf{x}_{r,\mathcal{I}}^{(m_i^*)}(t), \hat{\mathbf{x}}_{r,\mathcal{J}}(t), \boldsymbol{\mu}(t), \boldsymbol{\theta}_k^{(m_k)})$ is a function of five arguments: (i) the k th extent that is being modeled, $x_{r,k}^{(m_k)}(t)$, (ii) the $k - 1$ extents that have been modeled, $\mathbf{x}_{r,\mathcal{I}}^{(m_i^*)}(t)$ with $\mathcal{I} = \{1, 2, \dots, (k - 1)\}$, (iii) the $R - k$ extents that have not been modeled yet and are provided as measured and interpolated quantities, $\hat{\mathbf{x}}_{r,\mathcal{J}}(t)$ with $\mathcal{J} = \{k + 1, \dots, R\}$, (iv) the extents of inlet and of initial conditions, $\boldsymbol{\mu}(t)$, and (v) the adjustable model parameters of the k th reaction, $\boldsymbol{\theta}_k^{(m_k)}$. Note that, in the formulation of Eq. (5.8), the parameters of the first k models are estimated simultaneously. Also, since the number of provided measured extents ($R - k$) decreases when k increases, the interpolation bias is reduced progressively over the identification steps.

5.1.4 Bias due to error propagation and interpolation

As discussed in [16], the extent-based identification methods, whether incremental or sequential (except for the last step when $\mathcal{J} = \emptyset$), use measured quantities to model the extents. This introduces a bias for two reasons: (i) the propagation of measurement errors in nonlinear models, and (ii) the interpolation of quantities measured at a few discrete time instants $t_h, \forall h = 1, \dots, H$, to a dense number of time points $t_p \in [t_1, t_H]$ for the purpose of integration. These two cases are discussed next.

Structural Bias: The structural bias results from the propagation of measurement errors through the nonlinear kinetic model. To illustrate this bias, consider the kinetic model $\varphi_n(t) = k \left(\frac{n_A(t)}{V(t)} \right) \left(\frac{n_B(t)}{V(t)} \right)$. Assume that the volume is perfectly known, $\tilde{V}(t_h) = V_{true}(t_h)$, but the reconciled numbers of moles $\hat{n}_A(t_h)$ and $\hat{n}_B(t_h)$ are corrupted by additive and zero-mean Gaussian noise of known variance $\sigma_{\hat{n}_A}^2, \sigma_{\hat{n}_B}^2$ and $cov(\hat{n}_A, \hat{n}_B)$. Defining $k' = \frac{k}{\tilde{V}(t)^2}$, the expected value of φ_n at time t_h using measured quantities is:

$$\begin{aligned} E[\hat{\varphi}_n(t_h)] &= k' E[\hat{n}_A(t_h) \hat{n}_B(t_h)] \\ &= k' E[\hat{n}_A(t_h)] E[\hat{n}_B(t_h)] + k' cov(\hat{n}_A, \hat{n}_B) \end{aligned} \quad (5.9)$$

$$= k' \underbrace{n_{A,true}(t_h) n_{B,true}(t_h)}_{\bar{\varphi}_n(t_h)} + k' \underbrace{cov(\hat{n}_A, \hat{n}_B)(t_h)}_{\bar{b}_n(t_h)}, \quad (5.10)$$

which demonstrates that plugging measured quantities in the kinetic model φ_n leads to a biased estimated value, with the statistical bias $\bar{b}_n := k' cov(\hat{n}_A, \hat{n}_B)(t_h)$. For a more accurate model identification and parameter estimation, this bias must be accounted for. The unbiased model estimate of φ_n , denoted $\bar{\varphi}_n$, is obtained by rearranging Eq. (5.10) so that $\bar{\varphi}_n(t_h) = E[\hat{\varphi}_n(t_h)] - \bar{b}_n(t_h)$. In practice, for a given noise realization, a bias-corrected model estimate of $\varphi_n(t_h)$ is obtained by removing the bias $\hat{b}_n(t_h)$ from

Algorithm 3 Extent-based sequential model identification (SMI)

Step 1: Compute $\mathbf{x}_{in}(t)$, $\mathbf{x}_{ic}(t)$ and $\hat{\mathbf{n}}^{vRV}(t)$ using Eqs. (3.10b), (3.10c) and (3.29)

Step 2: Compute $\hat{\mathbf{x}}_r(t)$ for all R reactions using Eq. (3.32)

Step 3: Sort the computed extents in $\hat{\mathbf{x}}_r$ based on the order in which they will be modeled.

Step 4: The model identification task is carried out by postulating rate expressions for each reaction sequentially (one after the other). Let the subscript k denote the reaction being modeled, \mathcal{I} the set of reactions that have been modeled and \mathcal{J} the set of reactions that will be modeled later. The model identification problem for the R reactions can be formulated as follows:

for $k = 1, \dots, R$

$\mathcal{I} = \{1, 2, \dots, k-1\}$

$\mathcal{J} = \{k+1, k+2, \dots, R\}$

for $m_k = 1, \dots, M_k$

$$\begin{aligned} \begin{bmatrix} \theta_{\mathcal{I}}^{(m_{\mathcal{I}}^*)} \\ \theta_k^{(m_k^*)} \end{bmatrix} &= \arg \min_{\theta_{\mathcal{I}}^{(m_{\mathcal{I}}^*)}, \theta_k^{(m_k^*)}} J(\theta_{\mathcal{I}}^{(m_{\mathcal{I}}^*)}, \theta_k^{(m_k^*)}) \\ \text{s.t. } J &= \sum_{h=1}^H \left(\begin{bmatrix} \hat{\mathbf{x}}_{r,\mathcal{I}}(t_h) \\ \hat{x}_{r,k}(t_h) \end{bmatrix} - \begin{bmatrix} \mathbf{x}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t_h) \\ x_{r,k}^{(m_k^*)}(t_h) \end{bmatrix} \right)^T \mathbf{W}_{1:k}(t_h) \left(\begin{bmatrix} \hat{\mathbf{x}}_{r,\mathcal{I}}(t_h) \\ \hat{x}_{r,k}(t_h) \end{bmatrix} - \begin{bmatrix} \mathbf{x}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t_h) \\ x_{r,k}^{(m_k^*)}(t_h) \end{bmatrix} \right) \\ \dot{\mathbf{x}}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t) &= V(t) \varphi_{x,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(\mathbf{x}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t), x_{r,k}^{(m_k^*)}(t), \hat{\mathbf{x}}_{r,\mathcal{J}}(t), \boldsymbol{\mu}(t), \theta_{\mathcal{I}}^{(m_{\mathcal{I}}^*)}) - \omega(t) \mathbf{x}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t), \\ &\quad \mathbf{x}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(0) = \mathbf{0}_{k-1} \\ \dot{x}_{r,k}^{(m_k^*)}(t) &= V(t) \varphi_{x,k}^{(m_k^*)}(\mathbf{x}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t), x_{r,k}^{(m_k^*)}(t), \hat{\mathbf{x}}_{r,\mathcal{J}}(t), \boldsymbol{\mu}(t), \theta_k^{(m_k^*)}) - \omega(t) x_{r,k}^{(m_k^*)}(t), \\ &\quad x_{r,k}^{(m_k^*)}(0) = 0. \end{aligned} \quad (5.8)$$

end for

$$m_k^* = \arg \min_{m_k} \left\{ J(\theta_{\mathcal{I}}^{(m_{\mathcal{I}}^*)}, \theta_k^{(m_k^*)}) \right\}_{m_k=1}^{M_k}$$

end for

with $\mathbf{W}_{1:k}(t_h)$ the inverse of the first k -dimensional submatrix of the variance-covariance matrix $\boldsymbol{\Sigma}_x$ at time t_h , and the vector quantities $\mathbf{x}_{r,\mathcal{I}}^{(m_{\mathcal{I}}^*)} = [x_{r,1}^{(m_1^*)}, \dots, x_{r,k-1}^{(m_{k-1}^*)}]^T$, $\varphi_{\mathcal{I}}^{(m_{\mathcal{I}}^*)} = [\varphi_{x,1}^{(m_1^*)}, \dots, \varphi_{x,k-1}^{(m_{k-1}^*)}]^T$, $\hat{\mathbf{x}}_{r,\mathcal{J}} = [\hat{x}_{r,k+1}, \dots, \hat{x}_{r,R}]^T$ and $\theta_{\mathcal{I}}^{(m_{\mathcal{I}}^*)} = [\theta_1^{(m_1^*)}, \dots, \theta_{k-1}^{(m_{k-1}^*)}]^T$.

$\hat{\varphi}_n(t_h)$:

$$\bar{\varphi}_n(t_h) = \hat{\varphi}_n(t_h) - \hat{b}_n(t_h), \quad \forall h = 1, \dots, H. \quad (5.11)$$

For the kinetic model $\varphi_n(t) = k' n_A(t) n_B(t)$, a bias-corrected model estimate can be computed as

$$\bar{\varphi}_n(t_h) = k' \hat{n}_A(t_h) \hat{n}_B(t_h) - k' \text{cov}(\hat{n}_A, \hat{n}_B)(t_h), \quad \forall h = 1, \dots, H. \quad (5.12)$$

For a more accurate model identification and parameter estimation, the extent-based incremental and sequential approaches should be performed with $\bar{\varphi}_n$ instead of $\hat{\varphi}_n$. Note

that, for the sequential approach, the structural bias for the candidate models must be recomputed every time a new reaction is identified, since the number of measured signals decreases at each step.

Interpolation Bias: The sequential extent-based approach offers a remedy to that bias since the number of extents to be interpolated is reduced at every step and the need for interpolation finally disappears in the last step, which turns to be a simultaneous parameter estimation.

To illustrate this bias, consider the bias-corrected model of Eq. (5.12) and assume that the reconciled numbers of moles \hat{n}_A and \hat{n}_B are known at the discrete time instants t_h , $\forall h = 1, \dots, H$, but must be known for the purpose of integration at more frequent time points t_p , $\forall p = 1, \dots, P$, with $H < P$, $t_P = t_H$. Defining the interpolation errors on the numbers of moles of species A and B at time t_p as $\epsilon_A(t_p)$ and $\epsilon_B(t_p)$, the integral of $\bar{\varphi}_n$ between t_1 and t_H , defined as $\check{\Phi}_n(t_H)$, is:

$$\underbrace{\int_{t_1}^{t_H} \bar{\varphi}_n(\tau) d\tau}_{\check{\Phi}_n(t_H)} = \underbrace{\lim_{\Delta t_p \rightarrow 0} \sum_{p=1}^P \check{\varphi}_n(t_p) \Delta t_p}_{\bar{\Phi}_n(t_H)} + \underbrace{\lim_{\Delta t_p \rightarrow 0} \sum_{p=2}^P \check{b}_n(t_p) \Delta t_p}_{\bar{B}_n(t_H)}, \quad (5.13)$$

where Δt_p the time between two successive integration points. Note that, as the initial conditions are assumed to be perfectly known, the bias \check{b}_n at $t_1 = 0$ is zero and is thus omitted in the second summation term of Eq. (5.13).

For the example of the kinetic model $\varphi_n(t) = k' n_A(t) n_B(t)$, $\check{\varphi}_n(t_p)$ and $\check{b}_n(t_p)$ have the following structures:

$$\check{\varphi}_n(t_p) := k' \check{n}_A(t_p) \check{n}_B(t_p) - k' cov(\hat{n}_A, \hat{n}_B)(t_h) \quad (5.14a)$$

$$\check{b}_n(t_p) := k' (\check{n}_A(t_p) \epsilon_B(t_p) + \check{n}_B(t_p) \epsilon_A(t_p) + \epsilon_A(t_p) \epsilon_B(t_p)) \quad (5.14b)$$

where $\check{(\cdot)}$ indicate interpolated quantities. Eqs. (5.13) and (5.14) demonstrate that plugging interpolated quantities in the kinetic model $\bar{\varphi}_n$ (corrected for the structural bias according to Eq. (5.12)) for the purpose of integration leads to a biased estimated value $\check{\Phi}_n$, whose bias is \bar{B}_n . For an unbiased model identification and parameter estimation, this bias would need to be accounted for. Generally, an unbiased estimate of the integral of $\bar{\varphi}_n$ between t_1 and t_H would be:

$$\bar{\Phi}_n(t_H) = \check{\Phi}_n(t_H) - \bar{B}_n(t_H) \quad (5.15)$$

Unfortunately, as the interpolation errors $\epsilon_A(t_p)$ and $\epsilon_B(t_p)$ in Eq. (5.14b) required to compute the bias $\check{b}_n(t_p)$ are unknown, the interpolation bias cannot be removed.

5.2 Simulated Example

5.2.1 Homogeneous reaction system

5.2.1.1 Reaction system and operating conditions

Consider the acetoacetylation of pyrrole system given in Example 2.3 and Section 3.3.1. The reaction system is simulated in a semi-batch reactor with the following rate models for each of the four reactions:

$$r_1 = k_1 c_A c_B c_K \quad (5.16a)$$

$$r_2 = k_2 c_B^2 c_K \quad (5.16b)$$

$$r_3 = k_3 c_B \quad (5.16c)$$

$$r_4 = k_4 c_B c_C c_K. \quad (5.16d)$$

The reactor is operated as follows. The reactor initially contains 4 mol of A, 0.5 mol of B, 0.1 mol of C and 1 mol of catalyst K, for an initial volume of 0.41 L. Pure diketene (B) is fed into the reactor at the constant volumetric flowrate 0.1 L min^{-1} . The values of the rate constants are $k_1 = 0.0530 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$, $k_2 = 0.1280 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$, $k_3 = 0.0280 \text{ min}^{-1}$ and $k_4 = 0.0030 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$. Both the inlet flowrate and volume inside the reactor are assumed to be precisely known. The simulated numbers of moles and extents of reactions are shown in Fig. 5.1.

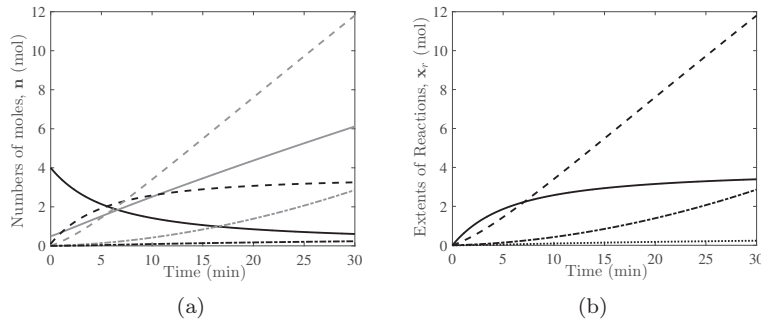


Figure 5.1 Acetoacetylation of pyrrole system in a semi-batch reactor: (a) Simulated numbers of moles of species A (—, black), B (—, gray), C (—, black), D (—, gray), E (---, gray), F (---, black), with species K constant at 1 mol (not shown) and (b) Simulated extents of reactions $x_{r,1}$ (—), $x_{r,2}$ (---), $x_{r,3}$ (---) and $x_{r,4}$ (---).

5.2.1.2 Simulated measurements and rate candidates

To obtain *measured* numbers of moles, the simulated numbers of moles of each species are corrupted by additive zero-mean Gaussian noise of standard deviation corresponding

to α % of its maximum value. The value of α is varied for the purpose of comparing the different model identification methods described in previous sections. The catalyst K is assumed to be known without any measurement errors.

In the following, the rate models and rate parameters of the four reactions are considered as unknown. Table 5.1 lists the set of rate candidates that will be tested for each reaction.

Table 5.1 Rate candidates for the acetoacetylation of pyrrole system.

R1	R2	R3	R4
$r_1^{(1)} = k_1 c_A c_B c_K$	$r_2^{(1)} = k_2 c_B^2 c_K$	$r_3^{(1)} = k_3 c_B$	$r_4^{(1)} = k_4 c_B c_C c_K$
$r_1^{(2)} = k_1 c_B$	$r_2^{(2)} = k_2 c_B$	$r_3^{(2)} = k_3 c_B^2$	$r_4^{(2)} = k_4 c_C$
$r_1^{(3)} = k_1 c_A$	$r_2^{(3)} = k_2 c_B^2$	$r_3^{(3)} = k_3 c_B c_K$	$r_4^{(3)} = k_4 c_B$
$r_1^{(4)} = k_1 c_K$	$r_2^{(4)} = k_2 c_B c_K$	$r_3^{(4)} = k_3 c_B^2 c_K$	$r_4^{(4)} = k_4 c_B c_C$
$r_1^{(5)} = k_1 c_A c_B$	$r_2^{(5)} = k_2 c_K$	$r_3^{(5)} = k_3 c_K$	$r_4^{(5)} = k_4 c_C c_K$
$r_1^{(6)} = k_1 c_A c_K$			
$r_1^{(7)} = k_1 c_B c_K$			
$r_1^{(8)} = k_1 c_A^2 c_K$			

5.2.1.3 Data sets

The performance of the three extent-based approaches presented are compared, namely, the incremental model identification formulated in terms of measured numbers of moles IMI_n , the incremental model identification formulated in terms of measured extents IMI_x , and the sequential model identification SMI. This comparison is performed using five different data sets, labeled D1 to D5. Data sets D1, D2 and D3 are generated by simulating the reaction system 1000 times with different noise realizations, 61 sampling points per species, and noise levels of 1%, 5% and 10%, respectively. Data set D4 is generated by simulating the system with 1000 different noise realizations, 21 sampling points per species and a noise level of 5%, while data set D5 is obtained by 1000 different noise realizations, 151 sampling points, and a noise level of 5%. Table 5.2 summarizes the noise level and the number of sampling points for each data set. Within each data set, the 1000 realizations are reconciled using the approach given in Chapter 4. The performance of all the model identification techniques is carried out without bias-correction.

Table 5.2 Noise level α and number of sampling points H for the different simulated data sets. Each data set corresponds to 1000 different noise realizations.

Data Set	Noise level (α)	Sampling points (H)
D1	1%	61
D2	5%	61
D3	10%	61
D4	5%	21
D5	5%	151

5.2.1.4 Comparison between IMI_n and IMI_x

The performance of the two extent-based incremental approaches IMI_n and IMI_x are compared for different noise levels and 61 sampling points. The results of the model identification are listed in Table 5.3. It is clearly shown that the variant IMI_x has a significantly higher probability of identifying the correct model than the variant IMI_n . Also, the standard deviations on the estimated parameters in the variant IMI_x is always smaller than the values computed by the variant IMI_n . These lower standard deviations can be explained by a reduced interpolation bias in the approach IMI_x compared to the approach IMI_n , since the values of the modeled extents in the approach IMI_x do not need to be interpolated (compare Eq. (5.4) and Eq. (5.6)). All these arguments lead to the conclusion that the extent-based incremental identification formulated in terms of measured extents (IMI_x) is superior to the same method formulated in terms of measured numbers of moles (IMI_n).

5.2.1.5 Comparison between IMI_x and SMI

Effect of the noise level

The performance of the extent-based incremental model identification formulated in terms of extents (variant IMI_x) is now compared to the performance of the sequential model identification (SMI), described in Section 5.1.3, for different noise levels and 61 sampling times. For SMI, the reaction steps are modeled in the following order based on the noise power: R2, R1, R3 and R4. The results of model identification by these two methods are presented in Table 5.4. It is clearly shown that SMI has a significantly higher probability of identifying the correct model than the IMI_x approach. If one excludes the identification of the first reaction in the modeling order, that is, reaction R2, which is treated in the same way in SMI and IMI_x , SMI always leads to better model identification (higher values of $\#_{/1000}$) and better parameter estimates (k^* closer to the true values and lower values of σ_{k^*}), once one or more rate models are fixed. For example, with 10% noise, the probability of finding the correct rate model for R1 is 73.1% for IMI_x , but 98.9% for SMI once the correct rate expression for R2 is fixed. Also, the standard deviations on the parameters (σ_{k^*}) reduce significantly over the steps of SMI, when more rate models are fixed. For example, with 10% noise, the standard

Table 5.3 Comparison between the extent-based incremental approaches IMI_n (Algorithm 1) and IMI_x (Algorithm 2) in terms of the number of correctly identified models over the 1000 realizations (labeled $\#_{/1000}$), as well as accuracy (k^*) and precision (σ_{k^*}) of the estimated parameters, using different noise levels and $H = 61$ sampling points. The true values of the rate constants are $k_1 = 0.053$, $k_2 = 0.128$, $k_3 = 0.028$ and $k_4 = 0.003$.

Reaction	k_{true}	Data set	α	IMI_n			IMI_x		
				$\#_{/1000}$	k^*	σ_{k^*}	$\#_{/1000}$	k^*	σ_{k^*}
R1	0.0530	D1	1%	995	0.0529	0.0009	1000	0.0530	0.0005
		D2	5%	733	0.0523	0.0041	942	0.0529	0.0023
		D3	10%	483	0.0519	0.0075	731	0.0530	0.0045
R2	0.1280	D1	1%	992	0.1275	0.0013	1000	0.1279	0.0007
		D2	5%	764	0.1250	0.0059	940	0.1271	0.0028
		D3	10%	425	0.1218	0.0114	924	0.1265	0.0059
R3	0.0280	D1	1%	983	0.0280	0.0001	984	0.0280	0.0001
		D2	5%	870	0.0279	0.0006	818	0.0279	0.0006
		D3	10%	833	0.0278	0.0011	756	0.0278	0.0010
R4	0.0030	D1	1%	749	0.0035	0.0032	999	0.0028	0.0001
		D2	5%	335	0.0038	0.0056	994	0.0028	0.0001
		D3	10%	236	0.0035	0.0059	866	0.0028	0.0002

deviation of k_2 estimated by SMI is reduced by more than 50% between the first step (R2 modeled with the other extents provided as measurements) and the last step (R4 modeled with all the other rate models fixed to their correct model structure), as shown in Table 5.4.

Fig. 5.2 gives a summary of the parameter estimates obtained by the IMI_n , IMI_x and SMI methods for all four reactions. It shows that the SMI method estimates rate parameters over a much more narrow region than IMI_x and IMI_n methods.

Table 5.4 Comparison between the extent-based incremental approaches in terms of extents (IMI_x) (Algorithm 2) and the sequential model identification (SMI) (Algorithm 3) in terms of the number of correctly identified models over the 1000 realizations (labeled #/1000), and accuracy (k^*) and precision (σ_{k^*}) of the estimated parameters for different noise levels and $H = 61$ sampling times. The true values of the rate constants are $k_2 = 0.128$, $k_1 = 0.053$, $k_3 = 0.028$ and $k_4 = 0.003$.

Reaction	Data set	α	IMI _x			SMI								
			#/1000	k^*	σ_{k^*}	#/1000	k_2^*	$\sigma_{k_2^*}$	k_1^*	$\sigma_{k_1^*}$	k_3^*	$\sigma_{k_3^*}$	k_4^*	$\sigma_{k_4^*}$
R2	D1	1%	1000	0.1279	0.0007	1000	0.1279	0.0007						
	D2	5%	940	0.1271	0.0028	940	0.1271	0.0028						
	D3	10%	924	0.1265	0.0059	924	0.1265	0.0059						
R1	D1	1%	1000	0.0530	0.0005	1000	0.1277	0.0007	0.0528	0.0002				
	D2	5%	942	0.0529	0.0023	999	0.1271	0.0022	0.0527	0.0009				
	D3	10%	731	0.0530	0.0045	989	0.1265	0.0043	0.0525	0.0018				
R3	D1	1%	984	0.0280	0.0001	1000	0.1274	0.0006	0.0528	0.0003	0.0280	0.0001		
	D2	5%	818	0.0279	0.0006	998	0.1273	0.0021	0.0528	0.0008	0.0280	0.0004		
	D3	10%	756	0.0278	0.0010	992	0.1274	0.0038	0.0528	0.0013	0.0281	0.0008		
R4	D1	1%	999	0.0028	0.00007	1000	0.1275	0.0003	0.0529	0.0001	0.0280	0.0001	0.0030	0.000007
	D2	5%	994	0.0028	0.0001	994	0.1274	0.0013	0.0528	0.0005	0.0280	0.0004	0.0030	0.000003
	D3	10%	866	0.0028	0.0002	895	0.1276	0.0025	0.0528	0.0009	0.0281	0.0008	0.0030	0.000007

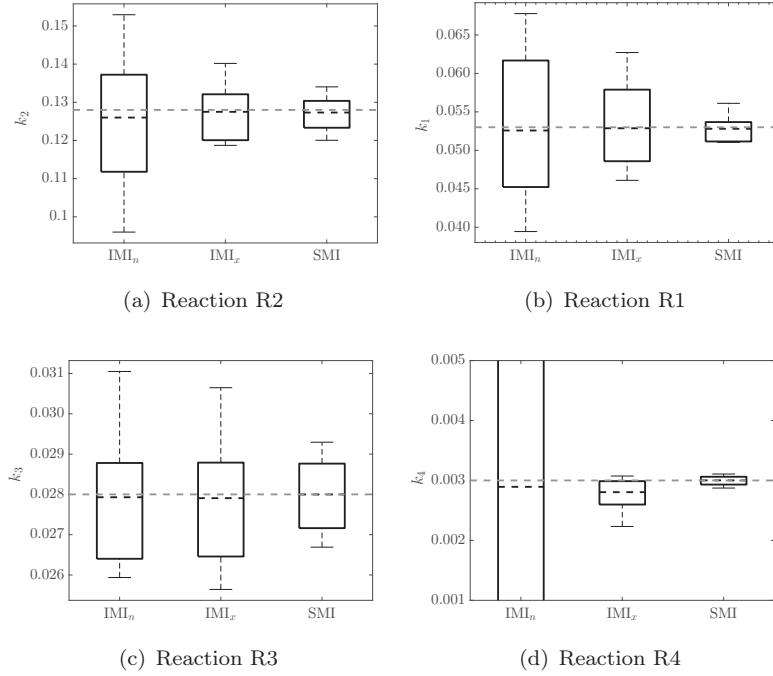


Figure 5.2 Summary of the parameter estimates with the true values (—, gray), minimum values (lowest error bar), 2.5 percentile values (bottom of box), median values (—, black), 97.5 percentile values (top of box) and maximum values (highest error bar) obtained by three different model identification methods for reactions (a) R2, (b) R1, (c) R3, and (d) R4. The minimum and 2.5 percentile values (0), the 97.5 percentile value (0.0220) and the maximum value (0.0221) are out of the bounds of Figure 5d for IMI_n .

Effect of the number of sampling points

In Table 5.5, the performance of IMI_x and SMI are compared for 5% noise but different numbers of sampling points. Note that the results for SMI in Table 5.5 correspond to the parameter values and standard deviation estimated for the last identification step, when R4 is modeled and all the other rate models are fixed to their correct model structure. Table 5.5 shows that SMI leads generally to smaller confidence intervals compared to IMI_x .

Discriminatory power

The average discriminatory power of the methods IMI_x and SMI are compared in Fig. 5.3 for the reaction steps R1, R3 and R4 using data with 5% noise and 21 sampling points¹. In this framework, the discriminatory power is defined as the ratio of the average sum of squares of a rate candidate to the average sum of squares of the correct rate model, such that the discriminatory power of the correct rate model is 1. Fig. 5.3

¹ The discriminatory power for R2 is not compared, as this reaction step is the first in the selection order of SMI and is hence modeled in the same way as in IMI_x .

Table 5.5 Comparison between the IMI_x and SMI methods in terms of the number of correctly identified models over the 1000 realizations (labeled $\#/\text{1000}$), and accuracy (k^*) and precision (σ_{k^*}) of the estimated parameters for different sampling times and a noise level of 5%.

Data set	H	Reaction	k_{true}	IMI_x			SMI		
				$\#/\text{1000}$	k^*	σ_{k^*}	$\#/\text{1000}$	k^*	σ_{k^*}
D4	21	R1	0.0530	808	0.0534	0.0036	984	0.0528	0.0009
		R2	0.1280	833	0.1276	0.0051	833	0.1274	0.0024
		R3	0.0280	738	0.0280	0.0009	994	0.0280	0.0007
		R4	0.0030	961	0.0028	0.0001	927	0.0030	0.0001
D2	61	R1	0.0530	942	0.0529	0.0023	999	0.0528	0.0005
		R2	0.1280	940	0.1271	0.0028	940	0.1274	0.0013
		R3	0.0280	818	0.0279	0.0006	998	0.0280	0.0004
		R4	0.0030	994	0.0028	0.0001	994	0.0030	0.00003
D5	151	R1	0.0530	997	0.0530	0.0015	1000	0.0528	0.0003
		R2	0.1280	940	0.1276	0.0021	940	0.1275	0.0009
		R3	0.0280	854	0.0279	0.0004	998	0.0280	0.0003
		R4	0.0030	993	0.0028	0.0001	999	0.0030	0.00002

shows that the discriminatory power of SMI is higher than that of IMI_x for R1, R3 and for some rate candidates of R4.

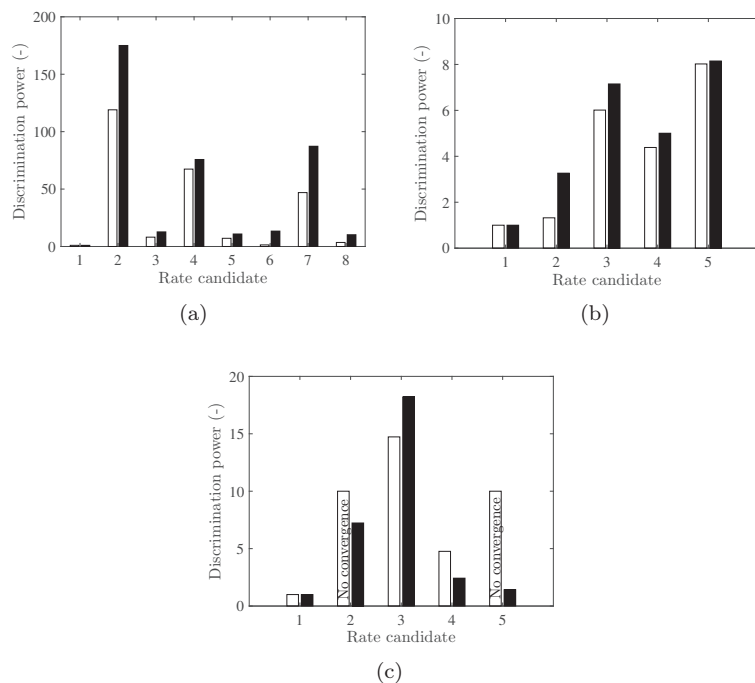


Figure 5.3 Discriminatory power (expressed as the mean sum of squares of a rate candidate normalized with respect to the mean sum of squares of the correct rate model) of the methods IMI_x (white bars) and SMI (black bars) for reaction (a) R1 and its 8 rate candidates (b) R3 and its 5 rate candidates, and (c) R4 and its 5 rate candidates. The *No convergence* label indicates that the parameter estimation did not converge and hence the corresponding rate candidate can certainly be excluded.

5.3 Summary

Identifying reliable kinetic models (and rate expressions) is one of the most challenging tasks in the modeling of chemical reaction systems. The extent-based incremental model identification approach has been introduced in the literature as an alternative to the computationally expensive simultaneous model identification. In the incremental approach, each rate process is modeled independently of each other. This chapter has provided an alternative model identification procedure for chemical reaction systems that combines the advantages of the incremental and simultaneous model identifications.

In the sequential model identification, the rate expressions for the different reaction steps of the system are identified one after the other. But, unlike the incremental approach, the rate expression that has been identified for a reaction step is then used to identify the rate expression for the next reaction step. Moreover, this chapter has shown that the bias due to interpolation is reduced at each step of the sequential identification

approach. The sequential model identification procedure has been illustrated via the simulated example of the acetoacetylation of pyrrole system.

Chapter 6

State Reconstruction and Estimation

Chapter 4 introduced the data reconciliation procedure in terms of the numbers of moles and the vessel extents for the case where all the species in a reaction system are measured. We assumed in Chapter 5 that the numbers of moles of all the species are available for kinetic model identification. While it is reasonable to expect these assumptions to be valid in the process development phase, only a subset of species are usually measured during process operation. The numbers of moles or concentrations of the unmeasured species must be inferred from the measured species. In this chapter, we propose two different methodologies for reconstructing the numbers of moles of the unmeasured species, while at the same time improving the accuracy of the measured species:

- *State reconstruction:* In this case, the conditions under which the numbers of moles of the unmeasured species can be reconstructed *without using kinetic information* are introduced.
- *State estimation:* In this case, the procedure of estimating the numbers of moles of the unmeasured species using a kinetic model and the measured numbers of moles are discussed.

For pedagogical reasons, this chapter discusses only the homogeneous reaction system as the corresponding extension to heterogeneous reaction system is straightforward. The chapter is organized as follows. Section 6.1.1 discusses the state reconstruction procedure using the formulation in terms of numbers of moles, while Section 6.1.2 deals with the corresponding procedure in terms of vessel extents. Section 6.2 compares the state estimation procedure in terms of both formulations.

6.1 State Reconstruction

In this section, we first discuss the state reconstruction procedure in terms of numbers of moles and in terms of vessel extents.

6.1.1 State reconstruction in terms of numbers of moles

Let $\tilde{\mathbf{n}}_a(t_h) = \mathbf{n}_{true,a}(t_h) + \epsilon_{n_a}$ denote the S_a -dimensional vector of noisy numbers of moles measured at the time instant t_h , $t_h \in [t_1, t_H]$, where ϵ_{n_a} is an S_a -dimensional vector of zero-mean Gaussian noise with the constant variance-covariance matrix Σ_{n_a} .

Assumption 6.1

Let \mathbf{N}^T , \mathbf{W}_{in} , \mathbf{n}_0 , $\mathbf{u}_{in}(t_h)$ and $u_{out}(t_h)$ be known without any errors.

If $S_a \geq d$, the numbers of moles $\tilde{\mathbf{n}}_u(t_h)$ of the S_u , unmeasured species can be reconstructed using $q_u = S_u$ invariant relationships as follows:

$$\tilde{\mathbf{n}}_u(t_h) = -(\mathbf{P}_u^T)^+ \mathbf{P}_a^T \tilde{\mathbf{n}}_a(t_h), \quad (6.1)$$

where, \mathbf{P}_a is the $S_a \times q_u$ submatrix of the invariant matrix \mathbf{P} corresponding to the S_a measured species, and \mathbf{P}_u is the $S_u \times q_u$ submatrix corresponding to the S_u unmeasured species.

Remark 6.1

If $S_a > d$, then $q_a = S_a - d$ invariants denoted \mathcal{I}_a are available for DR, and $q_u = q - q_a$ invariants denoted \mathcal{I}_u are used to reconstruct the unmeasured states using Eq. (6.1). In this case, the reconciliation problem can be formulated as follows:

$$\hat{\mathbf{n}}_a(t_{1:H}) = \arg \min_{\mathbf{n}_a(t_{1:H})} \sum_{i=1}^H (\tilde{\mathbf{n}}_a(t_i) - \mathbf{n}_a(t_i))^T \mathbf{W}_a(t_i) (\tilde{\mathbf{n}}_a(t_i) - \mathbf{n}_a(t_i)) \quad (6.2)$$

$$\begin{aligned} \text{s.t.} \quad & \mathcal{I}_a(\mathbf{n}_a(t_{1:H})) = \mathbf{0}_{q_a \times H} && (\text{Invariant constraints - measured states}) \\ & \mathcal{K}_{n,a}(\mathbf{n}_a(t_{1:H})) \leq \mathbf{0}_{k_{n,a} \times H} && (\text{Knowledge-based constraints - measured states}) \\ & \mathcal{K}_{n,u}(\mathbf{n}_u(t_{1:H})) \leq \mathbf{0}_{k_{n,u} \times H} && (\text{Knowledge-based constraints - unmeasured states}) \\ & \mathcal{M}_n(\mathbf{n}_a(t_{1:H})) \leq \mathbf{0}_{m_n \times H} && (\text{Measurement-based constraints}) \\ & \mathbf{n}(t_{1:H}) \geq \mathbf{0}_{S \times H}, && (\text{Positivity constraints}) \end{aligned}$$

with $\hat{\mathbf{n}}_a(t_{1:H}) = [\hat{\mathbf{n}}_a(t_1), \dots, \hat{\mathbf{n}}_a(t_H)]$ and $\hat{\mathbf{n}}_u(t_{1:H}) = [\hat{\mathbf{n}}_u(t_1), \dots, \hat{\mathbf{n}}_u(t_H)]$ the sequences of reconciled measured and unmeasured numbers of moles, respectively.

Remark 6.2

If $S_a = d$, then $q_a = 0$ and $q_u = q$, so that only the \mathcal{I}_u invariant relationships can be used for reconstructing the unmeasured states. Nevertheless, in this case also, the shape constraints can be used to smooth the S_a measured states.

Remark 6.3

If $S_a < d$, then the numbers of moles of the unmeasured species cannot be reconstructed using this approach.

6.1.2 State reconstruction in terms of vessel extents

Under assumption 6.1, let \mathbf{N}_a denote the $R \times S_a$ submatrix of \mathbf{N} corresponding to the S_a measured species. If \mathbf{N}_a has rank R , i.e. $S_a \geq R$, the numbers of moles of the remaining S_u species can be reconstructed without knowledge of reaction kinetics using the following steps:

1. Compute the vessel extent of inlets $\mathbf{x}_{in}(t_h)$ and initial conditions $x_{ic}(t_h)$ by integrating Eqs. (3.10b) and (3.10c):

$$\begin{aligned} \dot{\mathbf{x}}_{in}(t) &= \mathbf{u}_{in}(t) - \frac{u_{out}(t)}{m(t)} \mathbf{x}_{in}(t) & \mathbf{x}_{in}(0) &= \mathbf{0}_p \\ \dot{x}_{ic}(t) &= -\frac{u_{out}(t)}{m(t)} x_{ic}(t) & x_{ic}(0) &= 1 \\ \dot{m}(t) &= \mathbf{1}_p^T \mathbf{u}_{in}(t) - u_{out}(t) & m(0) &= m_0. \end{aligned} \quad (6.3)$$

2. Compute the numbers of moles of the measured species in the vessel reaction-variant form:

$$\tilde{\mathbf{n}}_a^{vRV}(t_h) = \tilde{\mathbf{n}}_a(t_h) - \mathbf{W}_{in,a} \mathbf{x}_{in}(t_h) - \mathbf{n}_{a,0} x_{ic}(t_h). \quad (6.4)$$

3. Compute the vessel extents of reactions using:

$$\tilde{\mathbf{x}}_r(t_h) = (\mathbf{N}_a^T)^+ \mathbf{n}_a^{vRV}(t_h). \quad (6.5)$$

4. The numbers of moles of the unmeasured species can be reconstructed as:

$$\tilde{\mathbf{n}}_u(t_h) = \mathbf{N}_u^T \tilde{\mathbf{x}}_r(t_h) + \mathbf{W}_{in,u} \mathbf{x}_{in}(t_h) + \mathbf{n}_{u,0} x_{ic}(t_h). \quad (6.6)$$

Remark 6.4

The vessel extents of reaction computed in Step 3 above can also be computed in a data-reconciliation formulation constrained by shape constraints as:

$$\hat{\mathbf{x}}_r(t_{1:H}) = \arg \min_{\mathbf{x}_r(t_{1:H})} \sum_{i=1}^H (\tilde{\mathbf{n}}_a^{vRV}(t_i) - \mathbf{n}_a^{vRV}(t_i))^T \mathbf{W}_a(t_i) (\tilde{\mathbf{n}}_a^{vRV}(t_i) - \mathbf{n}_a^{vRV}(t_i)) \quad (6.7)$$

$$\begin{aligned} s.t. \quad & \mathbf{n}_a^{vRV}(t_{1:H}) = \mathbf{N}_a^T \mathbf{x}_r(t_{1:H}) \\ & \mathcal{K}_x(\mathbf{x}_r(t_{1:H})) \leq \mathbf{0}_{k_x \times H} && \text{(Knowledge-based constraints)} \\ & \mathcal{M}_x(\mathbf{x}_r(t_{1:H})) \leq \mathbf{0}_{m_x \times H} && \text{(Measurement-based constraints)} \\ & \mathbf{n}_a(t_{1:H}) \geq \mathbf{0}_{S_a \times H} \text{ and } \mathbf{x}(t_{1:H}) \geq \mathbf{0}_{d \times H}. && \text{(Positivity constraints)} \end{aligned}$$

with $\hat{\mathbf{x}}_r(t_{1:H}) = [\hat{\mathbf{x}}_r(t_1), \dots, \hat{\mathbf{x}}_r(t_H)]$ the sequences of reconciled extents of reaction.

Asymptotical observer - Unknown initial conditions

Next, let us assume that there is significant uncertainty in the initial numbers of moles of the unmeasured species denoted by $\bar{\mathbf{n}}_{u,0}$. Let $\bar{\mathbf{n}}_u(t_h)$ and $\tilde{\mathbf{n}}_u(t_h)$ denote the estimated number of moles with the uncertain initial conditions and the estimated numbers of moles with the correct initial conditions

Proposition 6.1

The estimation error $\mathbf{e}(t_h) := \bar{\mathbf{n}}_u(t_h) - \tilde{\mathbf{n}}_u(t_h)$ converges asymptotically to zero in the presence of an outlet flow.

Proof: $(\mathbf{N}_a^T)^+$ represents the Moore-Penrose pseudo-inverse of \mathbf{N}_a^T . This pseudo-inverse exists and is unique if \mathbf{N}_a has rank R . It follows that the estimation error $\mathbf{e}(t_h) := \bar{\mathbf{n}}_u(t_h) - \tilde{\mathbf{n}}_u(t_h)$ can be evaluated from (6.6) as:

$$\mathbf{e}(t_h) = x_{ic}(t_h) \mathbf{e}(0). \quad (6.8)$$

Computing the time derivative of the error and using the expression in Eq. (3.10c) for $\dot{x}_{ic}(t_h)$ gives:

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \dot{x}_{ic}(t) \mathbf{e}(0) = -\frac{u_{out}(t)}{m(t)} x_{ic}(t) \mathbf{e}(0) \\ &= -\frac{u_{out}(t)}{m(t)} \mathbf{e}(t) \quad \mathbf{e}(0) = \bar{\mathbf{n}}_u(0) - \mathbf{n}_{u,0}. \end{aligned} \quad (6.9)$$

Hence, the estimation error goes asymptotically to zero if $u_{out}(t) \neq 0$. \square

6.2 State Estimation

In this part of the chapter, we give the procedure for improving the quality of measured signals and reconstructing the unmeasured signals *using kinetic information*. We show that the knowledge based constraints (\mathcal{K}) developed in Chapter 4 in terms of numbers of moles and in terms of vessel extents can be used to enhance the quality of the state estimates compared to the estimation without these constraints. First, we give a procedure for identifying measurement-based shape constraints using measurements and kinetic models. For the sake of simplicity, this procedure is presented in the context of the extent domain, but it can also be applied in the concentration domain. The procedure is as follows:

6.2.1 Measurement-based constraints

1. Using Eq. (3.10) and noting that \mathbf{r}_v is a function of \mathbf{x} , where $\mathbf{x} = \begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{in}(t) \\ x_{ic}(t) \end{bmatrix}$, express the first and second time derivatives of the extents *analytically* in terms of \mathbf{x} , \mathbf{u}_{in} and ω .
2. Select a time window \mathcal{T} of size N .
3. Compute the extents $\tilde{\mathbf{x}}(t_h) = \mathcal{T} \tilde{\mathbf{n}}(t_h)$ in the time window \mathcal{T} from the *measured* numbers of moles $\tilde{\mathbf{n}}(t_h)$ ¹.
4. Estimate the first and second derivatives of each extent using the analytical expressions in Step (1), that is, $\dot{\tilde{x}}_i(t_h)$ and $\ddot{\tilde{x}}_i(t_h)$, $i = 1, \dots, d$, $t_h \in \mathcal{T}$.
5. Design shape constraints based on the sign of the estimated second derivatives:
 - if $\ddot{\tilde{x}}_i(t_h) \geq 0$, $\forall t_h \in \mathcal{T}$, then $\tilde{x}_i(t)$ is convex on \mathcal{T} , and stop the procedure;
 - if $\ddot{\tilde{x}}_i(t_h) \leq 0$, $\forall t_h \in \mathcal{T}$, then $\tilde{x}_i(t)$ is concave on \mathcal{T} , and stop the procedure;
 - if $\ddot{\tilde{x}}_i(t_h)$ changes sign on the time window \mathcal{T} , reduce the size of the time window and go back to Step (2); however, if N is already small with regard to the window size that is necessary to handle measurement noise, no convex/concave shape can be imposed and proceed to Step (6) to investigate the existence of monotonicity constraints.
6. Design shape constraints based on the sign of the estimated first derivatives:
 - if $\dot{\tilde{x}}_i(t_h) \geq 0$, $\forall t_h \in \mathcal{T}$, then $\tilde{x}_i(t)$ is monotonically increasing on \mathcal{T} , and stop the procedure;
 - if $\dot{\tilde{x}}_i(t_h) \leq 0$, $\forall t_h \in \mathcal{T}$, $\tilde{x}_i(t)$ is monotonically decreasing on \mathcal{T} , and stop the procedure;
 - in case $\dot{\tilde{x}}_i(t_h)$ changes sign on the time window \mathcal{T} , there are no observable shape constraints; however, the nonnegative properties of $\tilde{\mathbf{x}}_{in}(t)$, $\tilde{x}_{ic}(t)$, and of $\tilde{\mathbf{x}}_r(t)$ for irreversible reactions remain valid.

6.2.2 System representation

For state estimation in a stochastic framework, it is necessary to extend the system representations in terms of numbers of moles given in Section 2.1.2 and in terms of

¹ The extents can also be computed using the data reconciliation procedure described in Chapter 4.

vessel extents given in Section 3.1.3 with measurement equations and both process and measurement noises.

System representation: Numbers of moles

Since the dimensionality of the reaction system is d and not S , it is possible to use the invariant relationships to rewrite Eq. (2.6) in terms of d independent species. The dynamic model can then be rewritten as:

$$\dot{\mathbf{n}}_1(t) = \mathbf{N}_1^T \mathbf{r}_v(t) + \mathbf{W}_{in,1} \mathbf{u}_{in}(t) - \omega(t) \mathbf{n}_1(t), \quad \mathbf{n}_1(0) = \mathbf{n}_{01} \quad (6.10a)$$

$$\mathbf{n}_2(t) = -(\mathbf{P}_2) \mathbf{P}_1^+ \mathbf{n}_1(t), \quad (6.10b)$$

where \mathbf{n}_1 is the d -dimensional vector of independent species, \mathbf{n}_2 the q -dimensional vector of dependent species, \mathbf{N}_1 is the $R \times d$ subset of the stoichiometric matrix, $\mathbf{W}_{in,1}$ the $d \times p$ subset of inlet compositions, \mathbf{n}_{01} the d -dimensional vector of initial conditions, \mathbf{P}_2 is the $q \times q$ subset of \mathbf{P} corresponding to the dependent species and \mathbf{P}_1 the $d \times q$ subset of \mathbf{P} corresponding to the independent species. Note that the set of independent species are chosen such that rank of the matrix $[\mathbf{N}_1^T \ \mathbf{W}_{in,1} \ \mathbf{n}_{01}] = d$. The system can be written in an stochastic form as:

$$\dot{\mathbf{n}}_1(t) = \mathbf{N}_1^T \mathbf{r}_v(t) + \mathbf{W}_{in,1} \mathbf{u}_{in}(t) - \omega(t) \mathbf{n}_1(t) + \mathbf{w}_{n1}(t), \quad \mathbf{n}_1(0) = \mathbf{n}_{01} \quad (6.11a)$$

$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{I}_d \\ -(\mathbf{P}_2) \mathbf{P}_1^+ \end{bmatrix} \mathbf{n}_1(t) + \mathbf{w}_y(t) \quad (6.11b)$$

The S -dimensional measurement vector \mathbf{y} contains the measured numbers of moles. The term \mathbf{w}_y represents Gaussian white measurement noise of covariance \mathbf{Q}_y . The term \mathbf{w}_{n1} is a Gaussian random variable with zero-mean and constant variance-covariance matrix \mathbf{Q}_n . Note, in the previous chapters, the variance-covariance matrix was denoted by Σ_n .

System representation: Vessel extents

For System (3.10), one can write:

$$\dot{\mathbf{x}}_r(t) = \mathbf{r}_v(t) - \omega(t) \mathbf{x}_r(t) + \mathbf{w}_r(t), \quad \mathbf{x}_r(0) = \mathbf{0}_R \quad (6.12a)$$

$$\dot{\mathbf{x}}_{in}(t) = \mathbf{u}_{in}(t) - \omega(t) \mathbf{x}_{in}(t) + \mathbf{w}_{in}(t), \quad \mathbf{x}_{in}(0) = \mathbf{0}_p \quad (6.12b)$$

$$\dot{x}_{ic}(t) = -\omega(t) x_{ic}(t) + w_{ic}(t), \quad x_{ic}(0) = 1 \quad (6.12c)$$

$$\mathbf{y}(t) = \mathbf{N}^T \mathbf{x}_r(t) + \mathbf{W}_{in} \mathbf{x}_{in}(t) + \mathbf{n}_0 x_{ic}(t) + \mathbf{w}_y(t). \quad (6.12d)$$

The terms \mathbf{w}_r , \mathbf{w}_{in} and w_{ic} are Gaussian random variables with zero-mean and constant variance-covariance \mathbf{Q}_r , \mathbf{Q}_{in} , and q_{ic} , respectively.

6.2.3 Receding-horizon nonlinear Kalman filter

In this section, the RNK filter equations are developed for the formulation in terms of extents. The corresponding equations in terms of numbers of moles can be written similarly and are not discussed here. In Section 6.2.3.1, the RNK filter equations are derived purely for a state estimation problem with no model adjustment, while Section 6.2.3.2, discusses simultaneous state estimation and model adjustment.

6.2.3.1 State estimation

Here, it is assumed that Assumption 6.1 is valid. For ease of notation, the right-hand sides of Eqs. (6.12a)–(6.12c) are defined as $\mathbf{f}_r(\cdot)$, $\mathbf{f}_{in}(\cdot)$ and $f_{ic}(\cdot)$ and are aggregated to the d -dimensional vector $\mathbf{f}_x(\cdot)$; similarly, the block-diagonal covariance matrix \mathbf{Q}_x is formed:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_{in} \\ x_{ic} \end{bmatrix}, \mathbf{f}_x(\cdot) = \begin{bmatrix} \mathbf{f}_r(\cdot) \\ \mathbf{f}_{in}(\cdot) \\ f_{ic}(\cdot) \end{bmatrix}, \mathbf{Q}_x = \begin{bmatrix} \mathbf{Q}_r & 0 & 0 \\ 0 & \mathbf{Q}_{in} & 0 \\ 0 & 0 & q_{ic} \end{bmatrix}. \quad (6.13)$$

The RNK filter implements the prediction and update steps over a time window. Some important aspects of the RNK filter are briefly discussed next.

Prediction step

Given the state vector $\mathbf{x}(t_h|t_h)$, one computes the *a priori* estimate $\mathbf{x}(t_{h+1}|t_h), \dots, \mathbf{x}(t_{h+N}|t_h)$ for the time window \mathcal{T} of length N using the state evolution described by Eqs. (6.12a)–(6.12c). Let the (Nd) -dimensional vector $\mathbf{x}_{\mathcal{T}|t_h}$ contain all the concatenated states, i.e., $\mathbf{x}_{\mathcal{T}|t_h} := [\mathbf{x}(t_{h+1}|t_h)^\top, \dots, \mathbf{x}(t_{h+N}|t_h)^\top]^\top$. The prediction step is also called ‘open-loop’ estimation.

An *a priori* estimate of the covariance matrix $\mathbf{P}_{\mathcal{T}|t_h}$ of dimension $(Nd \times Nd)$ is given by

$$\mathbf{P}_{\mathcal{T}|t_h} = \begin{pmatrix} \mathbf{P}_{t_{h+1}|t_h} & \mathbf{P}_{(t_{h+1}t_{h+2})|t_h} & \cdots & \mathbf{P}_{(t_{h+1}t_{h+N})|t_h} \\ \mathbf{P}_{(t_{h+1}t_{h+2})|t_h} & \mathbf{P}_{t_{h+2}|t_h} & \cdots & \mathbf{P}_{(t_{h+2}t_{h+N})|t_h} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{(t_{h+1}t_{h+N})|t_h} & \cdots & \cdots & \mathbf{P}_{t_{h+N}|t_h} \end{pmatrix},$$

where the diagonal elements represent the variances of the predicted states and the off-diagonal elements represent the covariance between predicted states. The elements of the matrix $\mathbf{P}_{\mathcal{T}|t_h}$ are estimated from $\mathbf{P}(t_h|t_h)$ using the following iterative relationships [61]:

$$\begin{aligned} \mathbf{P}_{t_{h+N}|t_h} &= \mathbf{A}_{t_{h+N-1}}^\top \mathbf{P}_{t_{h+N-1}|t_h} \mathbf{A}_{t_{h+N-1}} + \mathbf{Q}_x \\ \mathbf{P}_{(t_{h+N-1})(t_{h+N})|t_h} &= \mathbf{P}_{(t_{h+N-1})(t_{h+N-1})|t_h} \mathbf{A}_{t_{h+N-1}}^\top. \end{aligned}$$

The recursion is initialized using

$$\mathbf{P}_{t_{h+1}|t_h} = \mathbf{A}_{t_h}^T \mathbf{P}_{t_h|t_h} \mathbf{A}_{t_h} + \mathbf{Q}_x,$$

where $\mathbf{A}_{t_h} := e^{\{\frac{\partial \mathbf{f}_x}{\partial \mathbf{x}}|_{\mathbf{x}(t_h|t_h)}\}}$ is the linearization matrix of the differential equations (6.12a)–(6.12c).

Update step

Given the N measured outputs $\mathbf{y}_{\mathcal{T}} := [\mathbf{y}(t_{h+1})^T, \dots, \mathbf{y}(t_{h+N})^T]^T$, the update step of RNK is formulated as a constrained optimization problem, whose solution is the *a posteriori* state estimate $\mathbf{x}_{\mathcal{T}|t_{h+N}} := [\mathbf{x}(t_{h+1}|t_{h+1})^T, \dots, \mathbf{x}(t_{h+N}|t_{h+N})^T]^T$. With the introduction of the quantities $\boldsymbol{\alpha} := \mathbf{x}_{\mathcal{T}|t_{h+N}} - \mathbf{x}_{\mathcal{T}|t_h}$ and $\boldsymbol{\beta} := \mathbf{y}_{\mathcal{T}} - \mathbf{f}_y(\mathbf{x}_{\mathcal{T}|t_h})$, the update step can be formulated as the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{x}_{\mathcal{T}|t_{h+N}}} \quad & \boldsymbol{\alpha}^T \mathbf{P}_{\mathcal{T}|t_h}^{-1} \boldsymbol{\alpha} + \boldsymbol{\beta}^T \mathbf{Q}_y^{-1} \boldsymbol{\beta} \\ \text{s.t.} \quad & \mathcal{K}(\mathbf{x}_{\mathcal{T}|t_{h+N}}) \leq \mathbf{0}_{k_x}, \\ & \mathcal{M}(\mathbf{x}_{\mathcal{T}|t_{h+N}}) \leq \mathbf{0}_{m_x}, \\ & \mathbf{x}_{\mathcal{T}|t_{h+N}} \geq \mathbf{0}_{Nd}, \end{aligned}$$

where $\mathcal{K}(\cdot)$ denotes the k_x applicable knowledge-based constraints and $\mathcal{M}(\cdot)$ denotes the m_x applicable measurement-based constraints.

The *a posteriori* covariance matrix $\mathbf{P}_{\mathcal{T}|t_{h+N}}$ is computed as follows [61]:

$$\mathbf{K}_{\mathcal{T}|t_{h+N}} = \mathbf{P}_{\mathcal{T}|t_h} \mathbf{C}_{\mathcal{T}|t_h} (\mathbf{C}_{\mathcal{T}|t_h} \mathbf{P}_{\mathcal{T}|t_h} \mathbf{C}_{\mathcal{T}|t_h}^T + \mathbf{Q}_n)^{-1} \quad (6.14)$$

$$\mathbf{P}_{\mathcal{T}|t_{h+N}} = (\mathbf{I} - \mathbf{K}_{\mathcal{T}|t_{h+N}} \mathbf{C}_{\mathcal{T}|t_h}) \mathbf{P}_{\mathcal{T}|t_h}, \quad (6.15)$$

where $\mathbf{C}_{\mathcal{T}|t_h}$ is the linearized measurement equation obtained at $\mathbf{x}_{\mathcal{T}|t_h}$. At the end of this prediction-update step, the scheme is repeated for the next time window of length N , that is, from t_{h+2} to t_{h+N+1} . Note that, in the absence of constraints, the RNK filter reduces to a traditional EKF filter.

6.2.3.2 Model adjustment

In the formulation of the state estimator in terms of vessel extents, the measurement equation $\mathbf{y}(t)$ is a function of the initial numbers of moles \mathbf{n}_0 . In practice, the initial numbers of moles \mathbf{n}_0 are almost always uncertain. Additionally, the model parameters ($\boldsymbol{\theta}$) estimated in the process development phase need not correspond to the actual parameter values during production. In such a case, the RNK problem can be formulated as a simultaneous state (extents) and parameter ($\mathbf{n}_0, \boldsymbol{\theta}$) estimation problem. Let $\mathbf{p} = [\mathbf{n}_0; \boldsymbol{\theta}]$. The parameters to be estimated are defined as pseudo-states with zero dynamics, i.e.,

$$\mathbf{p}(t_{h+1}) = \mathbf{p}(t_h) + \mathbf{Q}_p,$$

where \mathbf{Q}_p is the variance-covariance matrix of the parameters to be estimated. The new augmented vectors \mathbf{x} , $\mathbf{f}(\cdot)$ and the matrix \mathbf{Q}_x for this case can be defined as,

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_{in} \\ x_{ic} \\ \mathbf{p} \end{bmatrix}, \mathbf{f}_x(\cdot) = \begin{bmatrix} \mathbf{f}_r(\cdot) \\ \mathbf{f}_{in}(\cdot) \\ f_{ic}(\cdot) \\ \mathbf{0} \end{bmatrix}, \mathbf{Q}_x = \begin{bmatrix} \mathbf{Q}_r & 0 & 0 & 0 \\ 0 & \mathbf{Q}_{in} & 0 & 0 \\ 0 & 0 & q_{ic} & 0 \\ 0 & 0 & 0 & \mathbf{Q}_p \end{bmatrix}. \quad (6.16)$$

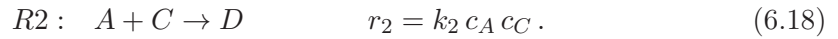
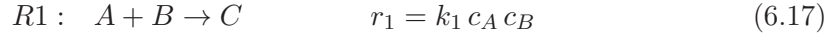
The prediction and update steps of the RNK filter can now be applied to this augmented system.

6.3 Simulated Example

This section uses a simulated example and RNK to compare the performance of constrained state estimation in terms of numbers of moles and in terms of extents. For simplicity, we do not illustrate the model adjustment procedure.

Reaction system

Consider the following two-reaction system,



The reaction system is simulated in a semi-batch reactor with $k_1 = 0.5$ and $k_2 = 0.3$, both in units $\text{L mol}^{-1} \text{ min}^{-1}$, $V = 1 \text{ L}$, $n_{A0} = 5 \text{ mol}$, and $n_{B0} = n_{C0} = 0 \text{ mol}$. The initial numbers of moles are assumed to be known. Species B is fed to the reactor with an inlet flowrate of 5 g min^{-1} . Species A , B and D are chosen as the independent species in the mole domain. The numbers of moles of species C are obtained using the invariant relation:

$$n_C(t) = n_{A0} + n_{C0} + 2n_{D0} - n_A(t) - 2n_D(t).$$

It is assumed that the numbers of moles of species A , B and D are measured every minute for 50 minutes. The simulated numbers of moles are corrupted with Gaussian white noise with the following measurement variance-covariance matrix:

$$\mathbf{Q}_y = \begin{bmatrix} 0.0806 & 0 & 0 \\ 0 & 0.0106 & 0 \\ 0 & 0 & 0.0553 \end{bmatrix}.$$

The flowrate and the volume are assumed to be perfectly known.

RNK filter in terms of numbers of moles

The differential equations are written as:

$$\dot{n}_A(t) = -\frac{\bar{k}_1}{V(t)}n_A(t)n_B(t) - \frac{\bar{k}_2}{V(t)}n_A(t)n_C(t) + w_{n,A} \quad (6.19a)$$

$$\dot{n}_B(t) = -\frac{\bar{k}_1}{V(t)}n_A(t)n_B(t) + w_{in,B}u_{in}(t) + w_{n,B} \quad (6.19b)$$

$$\dot{n}_D(t) = \frac{\bar{k}_2}{V(t)}n_A(t)n_C(t) + w_{n,C}, \quad (6.19c)$$

with the (incorrect) parameter values $\bar{k}_1 = 0.75$ and $\bar{k}_2 = 0.5$. The process noise matrix \mathbf{w}_n is assumed to be zero-mean and have the covariance matrix,

$$\mathbf{Q}_n = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.025 & 0 \\ 0 & 0 & 0.025 \end{bmatrix}.$$

The following constraints are known from prior knowledge:

- $n_A(t)$ is monotonically decreasing,
- $n_D(t)$ is monotonically increasing.

Furthermore, concave and convex constraints on all species are obtained from measurements using a window size $N = 10$.

RNK filter in terms of extents

The differential equations in the extent domain read:

$$\dot{x}_{r,1}(t) = \frac{\bar{k}_1}{V(t)}(n_{A0} - x_{r,1}(t) - x_{r,2}(t))(w_{in}x_{in}(t) - x_{r,1}(t)) + w_{x_{r,1}} \quad (6.20a)$$

$$\dot{x}_{r,2}(t) = \frac{\bar{k}_2}{V(t)}(n_{A0} - x_{r,1}(t) - x_{r,2}(t))(x_{r,1}(t) - x_{r,2}(t)) + w_{x_{r,2}} \quad (6.20b)$$

$$\dot{x}_{in}(t) = u_{in}(t) + w_{x_{in}}. \quad (6.20c)$$

The process noise matrix \mathbf{w}_x is zero-mean and has the covariance matrix \mathbf{Q}_x computed from \mathbf{Q}_n :

$$\mathbf{Q}_x = \begin{bmatrix} 0.125 & 0.025 & 0 \\ 0.025 & 0.025 & 0 \\ 0 & 0 & \epsilon \end{bmatrix},$$

with $\epsilon \rightarrow 0$ (the flowrate is perfectly known). For numerical reasons, ϵ is set to 10^{-5} .

The following constraints are known:

- $x_{r,1}(t)$ is concave,
- $x_{r,2}(t)$ is monotonically increasing,
- $x_{in}(t)$ is monotonically increasing.

Furthermore, concave and convex constraints on $x_{r,2}(t)$ and $\mathbf{x}_{in}(t)$ are obtained from measurements using the window size $N = 10$.

Results and Discussion

First, the performance of the two state estimators is compared using only constraints based on prior knowledge. The sum of squares of the errors between the true numbers of moles and the measured and estimated values are given in Table 6.1.

Table 6.1 Sum of squared errors for the unconstrained and estimated numbers of moles using an RNK filter formulated in numbers of moles (\mathbf{n}) and in extents (\mathbf{x}), and knowledge-based constraints.

Species	Unconstrained via \mathbf{n}	RNK estimation	
		via \mathbf{n}	via \mathbf{x}
A	0.96	0.44	0.10
B	0.19	0.13	0.06
C	1.98	0.63	0.27
D	0.52	0.21	0.12

Table 6.1 clearly shows that the addition of shape constraints improves the estimates. Furthermore, the performance is better in the extent domain than in the mole domain, which can be attributed to the stronger shape constraints that can be imposed in the extent domain.

Next, a similar comparison is done for the case where the shape constraints are determined from both prior knowledge and measurements. Table 6.2 lists the corresponding sum of squared errors. The corresponding simulated and estimated numbers of moles of A and D are shown in Fig. 6.1 for the formulation in terms of extents.

Table 6.3 shows that the performance of both estimators is improved in the presence of shape constraints obtained from measurements. The effectiveness of measurement-based constraints depends on the quality of the measured data, since the procedure relies on the computation of first and second derivatives of noisy measurements. At the limit, when the noise is too large, it might be impossible to apply shape constraints via measurements, and only the constraints from prior knowledge remain valid.

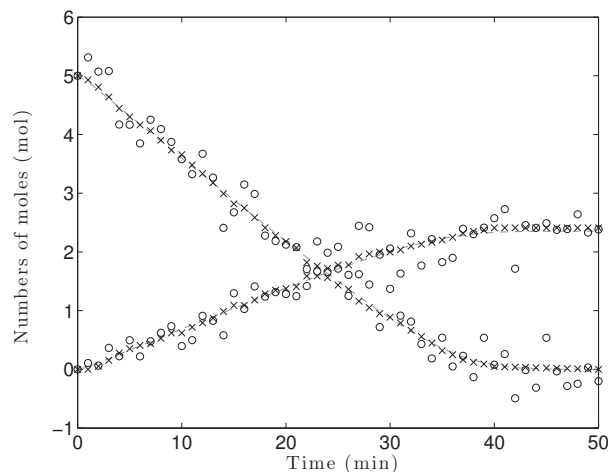


Figure 6.1 True (---), measured (\circ) and estimated number of moles obtained by RNK estimation via formulation in terms of extents for species A and D (\times) with constraints based on prior knowledge and measurements.

Table 6.2 Sum of squared errors for the measured and estimated numbers of moles using an RNK filter formulated in numbers of moles (\mathbf{n}) and in extents (\mathbf{x}), and both knowledge-based and measurement-based constraints.

Species	Unconstrained via \mathbf{n}	RNK estimation	
		via \mathbf{n}	via \mathbf{x}
A	0.96	0.27	0.06
B	0.19	0.07	0.04
C	1.98	0.37	0.26
D	0.52	0.13	0.10

6.4 Summary

This chapter has shown the procedure for reconstructing the numbers of moles of the unmeasured species without and with kinetic information. In the first part, a procedure for state reconstruction using the numbers of moles and vessel extents, in the absence of kinetic model, has been discussed. Next, the formulation of the state estimation problem in terms of numbers of moles and in terms of vessel extents have been compared. Furthermore, the addition of shape constraints to the state estimation problem has been shown to improve the accuracy of the estimated state variables. An alternate procedure for estimating these additional shape constraints from measurements has also been presented and demonstrated via a simulated example.

Chapter 7

Conclusions

7.1 Summary of Main Results

This section briefly summarizes the important results of each chapter in this dissertation along with some important questions that are worth investigating in the future.

Vessel Extents

A number of representations of chemical reaction systems exist in literature. In this thesis, we have shown that the representation of reaction systems in terms of vessel extents helps improve and speed up the process development phase compared to the traditional representation in terms of numbers of moles. At the same time, this representation gives a better understanding of the chemical reaction system as each state depends on a single rate process. A new linear transformation that converts the representation in terms of numbers of moles to the representation in terms of vessel extents has also been introduced. This dissertation has introduced the representation in terms of vessel extents for homogeneous reaction systems with heat exchange via jacket. This representation has also been extended to multi-phase reactions systems with mass transfer between phases.

Data Reconciliation

Data reconciliation uses redundancies expressed as relationships between state variables to reduce the noise in measured data. For chemical reaction systems, these relationships are derived from material and energy balance equations. This thesis has proposed a systematic procedure to derive these constraints for homogeneous and heterogeneous reaction systems. These relationships are algebraic constraints, since they do not contain information regarding past and future time instants.

This dissertation has also introduced the concept of shape constraints to provide dynamic information regarding the chemical reaction system in the absence of a kinetic model. We have shown how to identify *shape constraints*, such as monotonicity and curvature, on the numbers of moles and on vessel extents, which are constraints relating measurements over several time instants. Conditions under which these constraints are valid based on prior knowledge have been proposed. Furthermore, it has been shown that the alternative formulation of the reaction system in terms of vessel extents helps identify additional shape constraints.

Chapter 4 illustrates the advantages of using the identified shape constraints for data reconciliation. The higher number of constraints in the extent-based formulation leads to better reconciliation in this formulation compared to the formulation in terms of vessel extents. However, this dissertation has described knowledge-based constraints only for the case of irreversible reactions. Furthermore, the variance-covariance matrix of the reconciled estimates has been approximated using only equality constraints. Future work could focus on:

- Deriving knowledge-based constraints for systems with reversible reactions. Also, the impact of the inequality constraints on the variance-covariance matrix of the reconciled estimates could be investigated.

Sequential Kinetic Modelling

Identifying reliable descriptions of kinetics and mass transport is one of the most challenging tasks in the modeling of chemical reaction systems. This dissertation has introduced the extent-based sequential model identification approach, that combines the advantages of the incremental and the simultaneous approaches. Also, we have introduced a procedure to correct for the structural bias that could be introduced during the kinetic modeling step. Future work could focus on the following:

- Designing experiments using the sequential model identification. Since, at every step, a model is identified for each of the reactions, more specific experiments can be formulated in order to improve the discrimination for the reaction under study.

State Reconstruction and Estimation

This dissertation has compared the formulation of the state estimation problem in terms of numbers of moles and in terms of vessel extents. Further, the addition of shape constraints to the state estimation problem has also been shown to improve the accuracy of the estimated state variables. We have demonstrated that the formulation in terms of vessel extents gives better estimates of the states compared to the formulation in terms of numbers of moles. A procedure for estimating additional shape constraints from measurements has also been presented and demonstrated via a simulated example. Future work could focus on:

- The state estimation problem has been formulated in the Receding Horizon Nonlinear Kalman Filter framework. Further investigations must be made by comparing the performance of the two formulations in other frameworks such as the Moving Horizon Estimator. Also, the performance of the state estimator has been illustrated using simulated data; these methods should also be applied to process data in real time.

7.2 Outlook and Perspectives

The representation of chemical reaction systems in terms of vessel extents was introduced in Chapter 3. The advantages of this representation for data reconciliation, model identification and parameter estimation, as well as state estimation were investigated in Chapters 4, 5 and 6. This section focuses on some future applications of the representation in terms of vessel extents.

Fault Diagnosis

Process monitoring and fault diagnosis techniques are used for controlling quality and enforcing safety compliance in industrial processes [76, 77]. Processes are monitored by comparing abstract or physical variables to set points representing Normal Operating Conditions (NOC) and faults (AOC) are detected based on deviations from statistical thresholds. The representation in terms of vessel extents, where each state variable is a function of a single rate only, could provide a systematic way of identifying and isolating process faults. Fault diagnosis using extents can possibly be conducted either qualitatively or quantitatively.

- Qualitative approach: In this approach, the extents computed for the current batch can be compared with either the extents computed in previous batches or to the nominal values of the extents computed based on the process operation recipe. In the case where there is no fault in the system, then the trends of vessel extents should match with the expected trends. The approach using the operation recipe is especially useful, for example, in fine chemical industries, where the process recipe has to be followed exactly.
- Quantitative approach: In this approach, the extents computed from the current batch can be compared with the extents predicted using a process model. The most likely fault can then be estimated by using a statistical approach such as the Generalized Likelihood Approach (GLR) [54].

Control

Various control structures for continuous stirred-tank reactors based on reaction variants and extensive variables have been proposed in the literature [41, 78]. However, in the absence of a kinetic model, there does not exist a systematic way of tackling the control problem. One possible method is to compute the reaction rates from concentration and temperature measurements via the concept of vessel extents and then use the computed rates via a feedback linearization scheme for control [62].

For example, consider the temperature control problem in a homogeneous reactor with heat exchange. Let $Q_s(t)$ be the desired trajectory of the heat signal and $Q(t)$ be the measured heat signal. Let the heat exchanged between the jacket and the reactor, $q_{ex}(t)$, be the manipulated variable using the following control law:

$$q_{ex}(t) = v(t) - (-\Delta \mathbf{H})^T \hat{\mathbf{r}}_v(t) - \check{\mathbf{T}}_{in}^T \mathbf{u}_{in}(t) + \omega(t)Q(t), \quad (7.1)$$

$$v(t) = \dot{Q}_s(t) + \gamma(Q_s(t) - Q(t)), \quad (7.2)$$

where $\hat{\mathbf{r}}(t)$ represents the estimated reaction rates obtained by differentiating the vessel extents of reaction:

$$\hat{\mathbf{r}}_v(t) = \dot{\mathbf{x}}_r(t) + \omega(t)\mathbf{x}_r(t) \quad (7.3)$$

In this case, the feedback controller forces the control error $e(t) := Q_s(t) - Q(t)$ to converge exponentially to zero at the rate γ . This approach can then also be extended to handle the control of reactant and product concentrations, and simultaneous (multi-variable) control of temperature and concentrations.

Optimization

The possibility of reducing the duration of the iterations of static real-time optimization algorithms by applying the method of rate estimation and control using the concept of extents can also be investigated. In the context of real-time optimization of continuous processes, the estimated rates could be used to predict the economic outcome of the steady-state that corresponds to a given set of values of the decision variables, much before reaching that steady-state. Hence, this would allow fast convergence of continuous reactors to their optimal operating conditions.

Appendix A

Data Reconciliation - Heterogeneous Reaction Systems

In this chapter, the data reconciliation formulation for heterogeneous reaction systems is presented. For the sake of simplicity, we assume that $S_{f,a} = S_f$, i.e., we assume all the species in each phase is measured.

A.1 Reconciliation using Numbers of Moles

Data reconciliation in terms of numbers of moles is formulated as a weighted least-squares optimization problem constrained by the invariant relationships \mathcal{I}_f , the knowledge-based constraints \mathcal{K}_{n_f} , the measurement-based constraints \mathcal{M}_{n_f} , and positivity constraints. The reconciliation problem can be formulated as follows:

$$\hat{\mathbf{n}}_f(t_{1:H}) = \arg \min_{\mathbf{n}_f(t_{1:H})} \sum_{i=1}^H (\tilde{\mathbf{n}}_f(t_i) - \mathbf{n}_f(t_i))^T \mathbf{W}_f(t_i) (\tilde{\mathbf{n}}_f(t_i) - \mathbf{n}_f(t_i)) \quad (\text{A.1})$$

$$\begin{aligned} \text{s.t.} \quad & \mathcal{I}(\mathbf{n}_f(t_{1:H})) = \mathbf{0}_{q \times H} && (\text{Invariant constraints}) \\ & \mathcal{K}_{n_f}(\mathbf{n}_f(t_{1:H})) \leq \mathbf{0}_{k_{n_f} \times H} && (\text{Knowledge-based constraints}) \\ & \mathcal{M}_{n_f}(\mathbf{n}_f(t_{1:H})) \leq \mathbf{0}_{m_{n_f} \times H} && (\text{Measurement-based constraints}) \\ & \mathbf{n}_f(t_{1:H}) \geq \mathbf{0}_{S \times H}, && (\text{Positivity constraints}) \end{aligned}$$

with $\hat{\mathbf{n}}_f(t_{1:H}) = [\hat{\mathbf{n}}_f(t_1), \dots, \hat{\mathbf{n}}_f(t_H)]$ the $S_f \times H$ sequence of reconciled numbers of moles, $\hat{\mathbf{n}}_f(t_h)$ the vector of reconciled numbers of moles at time t_h , $h \in \{1, \dots, H\}$, and $\mathbf{W}_f(t_h) = \Sigma_{n_f}^{-1}(t_h)$ the weighting matrix at time t_h .

A.2 Reconciliation using Extents

Data reconciliation in terms of extents is formulated as a weighted least-squares optimization problem constrained by the knowledge-based \mathcal{K}_{x_f} , the measurement-based constraints \mathcal{M}_{x_f} , and positivity constraints. The reconciliation problem can be formulated as follows:

$$\hat{\mathbf{x}}_f(t_{1:H}) = \arg \min_{\mathbf{x}_f(t_{1:H})} \sum_{i=1}^H (\tilde{\mathbf{n}}_f(t_i) - \mathbf{n}_f(t_i))^T \mathbf{W}(t_i) (\tilde{\mathbf{n}}_f(t_i) - \mathbf{n}_f(t_i)) \quad (\text{A.2})$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{n}_f(t_{1:H}) = \mathbf{B} \mathbf{x}_f(t_{1:H}) \\ & \mathcal{K}_{x_f}(\mathbf{x}_f(t_{1:H})) \leq \mathbf{0}_{k_{x_f} \times H} && (\text{Knowledge-based constraints}) \\ & \mathcal{M}_{x_f}(\mathbf{x}_f(t_{1:H})) \leq \mathbf{0}_{m_{x_f} \times H} && (\text{Measurement-based constraints}) \\ & \mathbf{n}_f(t_{1:H}) \geq \mathbf{0}_{S \times H}, && (\text{Positivity constraints}) \end{aligned}$$

with $d_f \times H$ dimensional $\hat{\mathbf{x}}_f(t_{1:H}) := \begin{bmatrix} \hat{\mathbf{x}}_{r,f}(t_{1:H}) \\ \hat{\mathbf{x}}_{m,f}(t_{1:H}) \\ \hat{\mathbf{x}}_{in,f}(t_{1:H}) \\ \hat{\mathbf{x}}_{ic,f}(t_{1:H}) \end{bmatrix}$, and $\hat{\mathbf{x}}_{r,f}(t_{1:H}) = [\hat{\mathbf{x}}_{r,f}(t_1), \dots, \hat{\mathbf{x}}_{r,f}(t_H)]$, $\hat{\mathbf{x}}_{m,f}(t_{1:H}) = [\hat{\mathbf{x}}_{m,f}(t_1), \dots, \hat{\mathbf{x}}_{m,f}(t_H)]$, $\hat{\mathbf{x}}_{in,f}(t_{1:H}) = [\hat{\mathbf{x}}_{in,f}(t_1), \dots, \hat{\mathbf{x}}_{in,f}(t_H)]$ and $\hat{\mathbf{x}}_{ic,f}(t_{1:H}) = [\hat{x}_{ic,f}(t_1), \dots, \hat{x}_{ic,f}(t_H)]$ the sequences of reconciled extents of reaction, mass transfer, inlet and initial conditions, respectively. Note that the q_f invariant constraints are implicitly satisfied in this formulation since the invariants $\mathbf{x}_{iv,f}$ are set to zero (and hence absent) in $\hat{\mathbf{x}}_f(t_{1:H})$.

Appendix B

Sequential Model Identification - Heterogeneous reaction system

In this appendix, the sequential model identification procedure is extended to the heterogeneous reaction systems. In this case, models must be identified for the R_f reactions and p_m mass transfers in phase F. For the sake of generality, we define $\chi_f = \begin{bmatrix} \mathbf{x}_{r,f} \\ \mathbf{x}_{m,f} \end{bmatrix}$ as the $(R_f + p_m)$ -dimensional vector of the extents of reaction and mass transfer in phase F and its dynamics as:

$$\dot{\chi}_f(t) = \varphi_f(t) - \omega_f(t) \chi_f(t), \quad \chi(0) = \mathbf{0}_{R_f+p_m} \quad (\text{B.1})$$

with $\varphi_f := \begin{bmatrix} \mathbf{r}_{v,f} \\ \zeta_f \end{bmatrix}$ the $(R_f + p_m)$ -dimensional vector of rate expressions to model. Similarly, let us define the dynamics of $\mu_f = \begin{bmatrix} \mathbf{x}_{in,f} \\ x_{ic,f} \end{bmatrix}$ as:

$$\dot{\mu}_f(t) = \begin{bmatrix} \mathbf{u}_{in,f} \\ 0 \end{bmatrix} - \omega_f(t) \mu_f(t), \quad \mu_f(0) = \begin{bmatrix} \mathbf{0}_{p_f} \\ 1 \end{bmatrix} \quad (\text{B.2})$$

Also, let us introduce $\mathcal{M}_{n_f} = \{\mathcal{M}_{n_f,1}, \dots, \mathcal{M}_{n_f,R_f+p_m}\}$ the set of all $M = \sum_{k=1}^{R_f+p_m} M_k$ candidate models for the R reactions and p_m mass transfers expressed in terms of numbers of moles, and $\mathcal{M}_{\chi_f} = \{\mathcal{M}_{\chi_f,1}, \dots, \mathcal{M}_{\chi_f,R_f+p_m}\}$ the corresponding set in terms of extents.

B.1 Extent-based Sequential Model Identification (SMI)

The algorithm for the extent-based sequential model identification procedure is detailed in Algorithm 4. In Step 3, the vessel extents of reaction and extents of mass transfer are sorted based on the order in which they are to be modeled. The function $\varphi_{x,f,k}^{(m_k)}(\chi_{f,k}^{(m_k)}(t), \chi_{f,\mathcal{I}}^{(m_i^*)}(t), \hat{\chi}_{f,\mathcal{J}}(t), \mu_f(t), \theta_{f,k}^{(m_k)})$ is a function of five arguments: (i) the k th extent that is being modeled, $\chi_{f,k}^{(m_k)}(t)$, (ii) the $k - 1$ extents that have been modeled, $\chi_{f,i}^{(m_i^*)}(t)$, with $\mathcal{I} = \{1, 2, \dots, (k - 1)\}$, (iii) the $R_f + p_m - k$ extents that have not been modeled yet and are provided as interpolated quantities, $\hat{\chi}_{f,\mathcal{J}}(t)$, with $\mathcal{J} = \{k + 1, \dots, R_f + p_m\}$, (iv) the extents of inlet and of initial conditions, $\mu_f(t)$, and (v) the adjustable model parameters of the k th reaction, $\theta_{f,k}^{(m_k)}$. Note that, in the formulation of Eq. (5.8), the parameters of the first k models are estimated simultaneously.

Algorithm 4 Extent-based sequential model identification for heterogeneous system (SMI)

Step 1: Compute $\mathbf{x}_{in,f}(t)$, $\mathbf{x}_{ic,f}(t)$ and $\hat{\mathbf{n}}_f^{vRMV}(t)$ using Eqs. (3.39c), (3.39d) and (3.42)

Step 2: Compute $\hat{\chi}_f(t)$ for all the $R_f + p_m$ rate processes using Eq. (3.43)

Step 3: Sort the computed extents in $\hat{\chi}_f$ based on the order in which they will be modeled.

Step 4: The model identification task is carried out by postulating rate expressions for each rate process sequentially (one after the other). Let the subscript k denote the rate process being modeled, i the set of rate process that have been modeled and j the set of rate process that will be modeled later. The model identification problem for the $R_f + p_m$ reactions can be formulated as follows:

for $k = 1, \dots, R_f + p_m$

$i = \{1, 2, \dots, k-1\}$

$j = \{k+1, k+2, \dots, R_f + p_m\}$

for $m_k = 1, \dots, M_k$

$$\begin{aligned} \begin{bmatrix} \theta_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)} \\ \theta_{f,k}^{(m_k)^*} \end{bmatrix} &= \arg \min_{\theta_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}, \theta_{f,k}^{(m_k)^*}} J(\theta_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}, \theta_{f,k}^{(m_k)^*}) \\ \text{s.t. } J &= \sum_{h=1}^H \left(\begin{bmatrix} \hat{\chi}_{f,\mathcal{I}}(t_h) \\ \hat{\chi}_{f,k}(t_h) \end{bmatrix} - \begin{bmatrix} \chi_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t_h) \\ \chi_{f,k}^{(m_k)^*}(t_h) \end{bmatrix} \right)^T \mathbf{W}_{1:k,f}(t_h) \left(\begin{bmatrix} \hat{\chi}_{f,\mathcal{I}}(t_h) \\ \hat{\chi}_{f,k}(t_h) \end{bmatrix} - \begin{bmatrix} \chi_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t_h) \\ \chi_{f,k}^{(m_k)^*}(t_h) \end{bmatrix} \right) \\ \dot{\chi}_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t) &= V(t) \varphi_{\chi,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(\chi_{f,k}^{(m_k)^*}(t), \chi_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t), \hat{\chi}_{f,\mathcal{J}}(t), \boldsymbol{\mu}_f(t), \theta_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}) - \omega(t) \chi_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t), \\ &\quad \chi_{f,i}^{(m_i^*)}(0) = \mathbf{0}_{k-1} \\ \dot{\chi}_{f,k}^{(m_k)^*}(t) &= V(t) \varphi_{x,k}^{(m_k)}(\chi_{f,k}^{(m_k)^*}(t), \chi_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}(t), \hat{\chi}_{f,\mathcal{J}}(t), \boldsymbol{\mu}_f(t), \theta_{f,k}^{(m_k)^*}) - \omega(t) \chi_{f,k}^{(m_k)^*}(t), \\ &\quad \chi_{f,k}^{(m_k)^*}(0) = 0. \end{aligned} \quad (\text{B.3})$$

end for

$$m_k^* = \arg \min_{m_k} \left\{ J(\theta_{f,\mathcal{I}}^{(m_{\mathcal{I}}^*)}, \theta_{f,k}^{(m_k)^*}) \right\}_{m_k=1}^{M_k}$$

end for

with $\mathbf{W}_{1:k,f}(t_h)$ the inverse of the first k -dimensional submatrix of the variance-covariance matrix Σ_{χ_f} at time t_h , and the vector quantities $\chi_{f,\mathcal{I}}^{(m_i^*)} = [\chi_{f,1}^{(m_1^*)}, \dots, \chi_{f,k-1}^{(m_{k-1}^*)}]^T$, $\varphi_{\chi,\mathcal{I}}^{(m_i^*)} = [\varphi_{\chi,1}^{(m_1^*)}, \dots, \varphi_{\chi,k-1}^{(m_{k-1}^*)}]^T$, $\hat{\chi}_{f,\mathcal{J}} = [\hat{\chi}_{f,k+1}, \dots, \hat{\chi}_{f,R_f+p_m}]^T$ and $\theta_{f,\mathcal{I}}^{(m_i^*)} = [\theta_{f,1}^{(m_1^*)}, \dots, \theta_{f,k-1}^{(m_{k-1}^*)}]^T$.

Appendix C

Reaction Systems with Instantaneous Equilibria

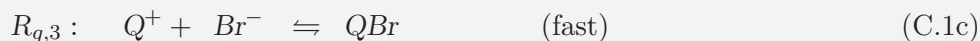
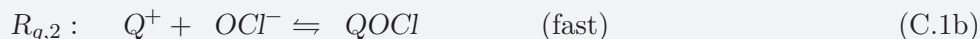
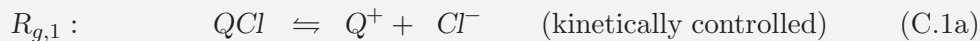
Consider that among the R_f reactions in phase F of a heterogeneous reaction system, $R_{f,k}$ are kinetically controlled reactions and $R_{f,e}$ are instantaneous equilibria, with $R_f = R_{f,k} + R_{f,e}$. Note that the subscripts $(\cdot)_k$ and $(\cdot)_e$ refer to kinetically controlled reactions and instantaneous equilibria, respectively.

Kinetic and equilibrium species: The set \mathcal{S}_f can be partitioned into the set $\mathcal{S}_{f,k}$ of kinetic species and the complementary set $\mathcal{S}_{f,e}$ of equilibrium species. The $\mathcal{S}_{f,k}$ kinetic species are only involved in the $R_{f,k}$ kinetically controlled reactions, that is, not in the instantaneous equilibria. The $\mathcal{S}_{f,e}$ equilibrium species are involved in the $R_{f,e}$ instantaneous equilibria (and possibly in some of the kinetically controlled reactions). Hence, $\mathcal{S}_f = \mathcal{S}_{f,k} \cup \mathcal{S}_{f,e}$ and $S_f = S_{f,k} + S_{f,e}$.

Equilibrium components: An equilibrium component is a molecule constituent that is conserved by instantaneous equilibria. The set $\mathcal{S}_{f,c}$ contains the $S_{f,c}$ equilibrium components that are found in $\mathcal{S}_{f,e}$ and whose total numbers of moles are conserved in the instantaneous equilibria. The subscript $(\cdot)_c$ in this appendix will be used to indicate an equilibrium component that is conserved by the instantaneous equilibria. The reduction of the $\mathcal{S}_{f,e}$ equilibrium species to the $\mathcal{S}_{f,c}$ equilibrium components is a linear operation that can be represented by the matrix \mathbf{E}_f of dimension $S_{f,c} \times S_{f,e}$ [45].

Example C.1 (Oxidation of Benzyl Alcohol with Hypochlorite)

To illustrate the notation, let us consider the instantaneous equilibrium reaction $Q^+ + OCl^- \rightleftharpoons QOCl$ that has two equilibrium components labeled Q_c (including Q^+ and $QOCl$) and OCl_c (including OCl^- and $QOCl$). In this simple example, with $\mathcal{S}_{f,e} = \{Q^+, OCl^-, QOCl\}$ and $\mathcal{S}_{f,c} = \{Q_c, OCl_c\}$, $\mathbf{E}_f = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. To continue with this example, consider the aqueous phase (denoted as phase G) of the heterogeneous oxidation of benzyl alcohol with hypochlorite:



This reaction system in phase G consists of one kinetically controlled reaction (Eq. C.1a) and two instantaneous equilibria (Eqs. C.1b and C.1c), thus giving $R_g = 3$, $R_{g,k} = 1$ and $R_{g,e} = 2$. The set of species in this phase is $\mathcal{S}_g = \{Cl^-, QCl, Q^+, OCl^-, Br^-, QOCl, QBr\}$, with the set of kinetic species $\mathcal{S}_{g,k} = \{Cl^-, QCl\}$ and the set of equilibrium species $\mathcal{S}_{g,e} = \{Q^+, OCl^-, Br^-, QOCl, QBr\}$. The set of components $\mathcal{S}_{g,c} = \{Q_c, OCl_c, Br_c\}$ can be constructed from $\mathcal{S}_{g,e}$ using the (3×5) matrix \mathbf{E}_g ,

$$\mathbf{E}_g = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

For a reaction system consisting of both kinetically controlled and equilibrium reactions, the balance equations for phase F can be written in two ways, namely, in terms of the set of all species \mathcal{S}_f as shown in section 2.2.1 or in terms of the kinetic species $\mathcal{S}_{f,k}$ and the equilibrium components $\mathcal{S}_{f,c}$, as discussed in the next subsection.

C.1 Balances for Kinetic Species and Equilibrium Components

The $S_f \times R_f$ stoichiometric matrix \mathbf{N}_f^T can be partitioned into two sub-matrices, the $S_{f,k} \times R_f$ matrix $\mathbf{N}_{f,k}^T$ associated with the $S_{f,k}$ kinetic species and the $S_{f,e} \times R_f$ matrix $\mathbf{N}_{f,e}^T$ associated with the $S_{f,e}$ equilibrium species:

$$\mathbf{N}_f^T = \begin{bmatrix} \mathbf{N}_{f,k}^T \\ \mathbf{N}_{f,e}^T \end{bmatrix}. \quad (C.2)$$

Similarly, the matrices $\mathbf{W}_{in,f}$ and $\mathbf{W}_{m,f}$ are partitioned into sub-matrices that are associated with the $S_{f,k}$ kinetic species, $\mathbf{W}_{in,f,k}$ and $\mathbf{W}_{m,f,k}$, and the $S_{f,e}$ equilibrium components, $\mathbf{W}_{in,f,e}$ and $\mathbf{W}_{m,f,e}$:

$$\mathbf{W}_{in,f} = \begin{bmatrix} \mathbf{W}_{in,f,k} \\ \mathbf{W}_{in,f,e} \end{bmatrix}, \quad \mathbf{W}_{m,f} = \begin{bmatrix} \mathbf{W}_{m,f,k} \\ \mathbf{W}_{m,f,e} \end{bmatrix}. \quad (C.3)$$

Considering that the numbers of moles of the equilibrium components can be written as $\mathbf{n}_{f,c}(t) = \mathbf{E}_f \mathbf{n}_{f,e}(t)$, the mole balance equations for the $S_{f,k}$ kinetic species and the $S_{f,c}$ components read:

$$\underbrace{\begin{bmatrix} \dot{\mathbf{n}}_{f,k}(t) \\ \dot{\mathbf{n}}_{f,c}(t) \end{bmatrix}}_{\dot{\mathbf{n}}_f(t)} = \underbrace{\begin{bmatrix} \mathbf{N}_{f,k}^T \\ \mathbf{E}_f \mathbf{N}_{f,e}^T \end{bmatrix}}_{\tilde{\mathbf{N}}_f^T} \mathbf{r}_{v,f}(t) + \underbrace{\begin{bmatrix} \mathbf{W}_{m,f,k} \\ \mathbf{E}_f \mathbf{W}_{m,f,e} \end{bmatrix}}_{\tilde{\mathbf{W}}_{m,f}} \zeta_f(t) + \underbrace{\begin{bmatrix} \mathbf{W}_{in,f,k} \\ \mathbf{E}_f \mathbf{W}_{in,f,e} \end{bmatrix}}_{\tilde{\mathbf{W}}_{in,f}} \mathbf{u}_{in,f}(t) - \omega_f(t) \underbrace{\begin{bmatrix} \mathbf{n}_{f,k}(t) \\ \mathbf{n}_{f,c}(t) \end{bmatrix}}_{\tilde{\mathbf{n}}_f(t)}, \quad (C.4)$$

where $\mathbf{n}_{f,k}$ is the $S_{f,k}$ -dimensional vector of numbers of moles of the kinetic species, $\mathbf{n}_{f,c}$ is the $S_{f,c}$ -dimensional vector of numbers of moles of the equilibrium components, and $(\bar{\cdot})$ indicates a quantity of dimension $\bar{S}_f := S_{f,k} + S_{f,c} \leq S_f$. The initial conditions for Eq. (C.4) are $\bar{\mathbf{n}}_f(0) = \begin{bmatrix} \mathbf{n}_{f,k}(0) \\ \mathbf{n}_{f,c}(0) \end{bmatrix} = \begin{bmatrix} \mathbf{n}_{f,k0} \\ \mathbf{n}_{f,c0} \end{bmatrix}$.

The matrix $\bar{\mathbf{N}}_f^T$ of dimension $\bar{S}_f \times R_f$ has a special structure with only zeros in the $R_{f,e}$ columns corresponding to the instantaneous equilibria. This follows from the way the kinetic species and the equilibrium components have been chosen, namely, (i) the $S_{f,k}$ kinetic species are only involved in the $R_{f,k}$ kinetically controlled reactions and not in the instantaneous equilibria, and (ii) the $S_{f,c}$ equilibrium components are conserved by the instantaneous equilibria and therefore all their corresponding stoichiometric coefficients are zero. It follows that $\bar{\mathbf{N}}_f^T$ has rank $R_{f,k} \leq R_f$, and the columns corresponding to the instantaneous equilibria can be discarded.¹ The resulting matrix $\bar{\mathbf{N}}_{f,k}^T$ has dimension $\bar{S}_f \times R_{f,k}$ and the vector $\mathbf{r}_{v,f}$ reduces to $\mathbf{r}_{v,f,k}$ of dimension $R_{f,k}$. The mole balance equation (C.4) becomes:

$$\dot{\bar{\mathbf{n}}}_f(t) = \bar{\mathbf{N}}_{f,k}^T \mathbf{r}_{v,f,k}(t) + \bar{\mathbf{W}}_{m,f} \boldsymbol{\zeta}_f(t) + \bar{\mathbf{W}}_{in,f} \mathbf{u}_{in,f}(t) - \omega_f(t) \bar{\mathbf{n}}_f(t), \quad \bar{\mathbf{n}}_f(0) = \bar{\mathbf{n}}_{f0}. \quad (\text{C.5})$$

Example C.1 (Oxidation of Benzyl Alcohol with Hypochlorite cont'd..)

To illustrate the matrix notation, consider the reaction system given by Eqs. (C.1a) - (C.1c), with $R_g = 3$, $R_{g,k} = 1$, $R_{g,e} = 2$, $S_g = 7$, $S_{g,k} = 2$, $S_{g,e} = 5$ and $S_{g,c} = 3$. The stoichiometric matrix \mathbf{N}_g^T with respect to all species (Cl^- , QCl , Q^+ , OCl^- , Br^- , QOCl , QBr) can be written as:

$$\mathbf{N}_g^T = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The corresponding stoichiometric matrix in terms of the kinetic species and the equilibrium components is:

¹ Because $R_{f,e}$ columns of $\bar{\mathbf{N}}_f^T$ are equal to zero, $S_{f,c} + R_{f,e} = S_{f,e}$.

$$\bar{\mathbf{N}}_g^T = \begin{bmatrix} \mathbf{N}_{g,k}^T \\ \mathbf{E}_g \mathbf{N}_{g,e}^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The last two columns of the matrix $\bar{\mathbf{N}}_g^T$ corresponding to the equilibrium reactions are all zeros and hence can be discarded, thus giving:

$$\bar{\mathbf{N}}_{g,k}^T = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

C.2 Reaction Systems with Equilibrium Reactions

For reaction systems with equilibrium reactions, two different cases can be considered for the system representation in terms of vessel extents .

1. In the first case, the extents are written in exactly the same way as the representation in Eq. (3.39). This includes extents for all reactions, that is there exists extents of reactions for both both kinetically controlled and instantaneous equilibrium reactions, in addition to extents of mass transfers, inlets and initial conditions. Note that the extents of equilibrium reactions capture the instantaneous shift in concentrations caused by the equilibria. The transformation from vessel extents to numbers of moles is performed using Eq. (3.40). The extents can be computed from the concentrations of “all species” present in the reaction system using the transformation described in Section 3.2.
2. In the second case, only the vessel extents of kinetically controlled reactions are considered for the system representation in terms of vessel extents. The transformation from vessel extents to numbers of moles produces the numbers of moles of the kinetic species and equilibrium components as shown in Eq. (C.7). The extents of the kinetically controlled reactions can be computed from the numbers of moles of the kinetic species and equilibrium components using the transformation shown in Section C.2.1. These species represent the minimal set that is necessary to describe both the kinetically controlled and the equilibrium reactions.

C.2.1 Using vessel extents of kinetic reactions

The representation in terms of vessel extents can be written as:

$$\begin{aligned}
 \dot{\mathbf{x}}_{r,f,k}(t) &= \mathbf{r}_{v,f,k}(t) - \omega_f(t) \mathbf{x}_{r,f,k}(t), & \mathbf{x}_{r,f,k}(0) &= \mathbf{0}_{R_{f,k}} \\
 \dot{\mathbf{x}}_{m,f}(t) &= \boldsymbol{\zeta}_f(t) - \omega_f(t) \mathbf{x}_{m,f}(t), & \mathbf{x}_{m,f}(0) &= \mathbf{0}_{p_m} \\
 \dot{\mathbf{x}}_{in,f}(t) &= \mathbf{u}_{in,f}(t) - \omega_f(t) \mathbf{x}_{in,f}(t), & \mathbf{x}_{in,f}(0) &= \mathbf{0}_{p_f} \\
 \dot{x}_{ic,f}(t) &= -\omega_f(t) x_{ic,f}(t), & x_{ic,f}(0) &= 1 \\
 \bar{\mathbf{x}}_{iv,f}(t) &= \mathbf{0}_{\bar{q}_f}.
 \end{aligned} \tag{C.6}$$

Transformation from \mathbf{x} to $\bar{\mathbf{n}}$: The numbers of moles of the kinetic species and equilibrium components can then be obtained as

$$\bar{\mathbf{n}}_f(t) = \bar{\mathbf{N}}_{f,k}^T \mathbf{x}_{r,f,k}(t) + \bar{\mathbf{W}}_{m,f} \mathbf{x}_{m,f}(t) + \bar{\mathbf{W}}_{in,f} \mathbf{x}_{in,f}(t) + \bar{\mathbf{n}}_{f0} x_{ic,f}(t). \tag{C.7}$$

Transformation from $\bar{\mathbf{n}}$ to \mathbf{x} : Let $\text{rank} \left([\bar{\mathbf{N}}_{f,k}^T \pm \bar{\mathbf{W}}_{m,f} \bar{\mathbf{W}}_{in,f} \bar{\mathbf{n}}_{f0}] \right) = \bar{d}_f$, with $\bar{d}_f := R_{f,k} + p_m + p_f + 1$, and the $\bar{S}_f \times \bar{q}_f$ matrix $\bar{\mathbf{P}}_f$ denote the null space of $[\bar{\mathbf{N}}_{f,k}^T \pm \bar{\mathbf{W}}_{m,f} \bar{\mathbf{W}}_{in,f} \bar{\mathbf{n}}_{f0}]^T$, with $\bar{q}_f := \bar{S}_f - \bar{d}_f$.² Then, the matrix $\bar{\mathcal{T}}_f = [\bar{\mathbf{N}}_{f,k}^T \bar{\mathbf{W}}_{m,f} \bar{\mathbf{W}}_{in,f} \bar{\mathbf{n}}_{f0} \bar{\mathbf{P}}_f]^{-1}$ transforms the \bar{S}_f -dimensional vector of numbers of moles $\bar{\mathbf{n}}_f$ of the kinetic species and equilibrium components into $R_{f,k}$ extents of kinetically controlled reaction $\mathbf{x}_{r,f,k}$, p_m extents of mass transfer $\mathbf{x}_{m,f}$, p_f extents of inlet $\mathbf{x}_{in,f}$, one extent of initial conditions $x_{ic,f}$ and \bar{q}_f invariants $\bar{\mathbf{x}}_{iv,f}$ that are identically equal to zero, as

$$\begin{bmatrix} \mathbf{x}_{r,f,k}(t) \\ \mathbf{x}_{m,f}(t) \\ \mathbf{x}_{in,f}(t) \\ x_{ic,f}(t) \\ \bar{\mathbf{x}}_{iv,f}(t) \end{bmatrix} = \bar{\mathcal{T}}_f \bar{\mathbf{n}}_f(t). \tag{C.8}$$

Compared to the transformation matrix \mathcal{T}_f , the matrix $\bar{\mathcal{T}}_f$ computes $R_{f,k}$ reaction extents, as it only extracts extents corresponding to the kinetically controlled reactions.

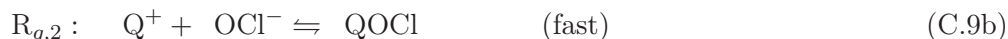
² $\bar{q}_f = q_f$ since $S_f = S_{f,k} + S_{f,e}$, $R_f = R_{f,k} + R_{f,e}$, and $S_{f,c} + R_{f,e} = S_{f,e}$.

C.3 Simulated Example

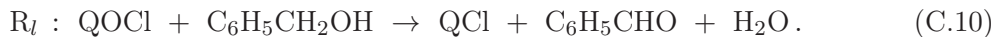
The vessel extent representation for reaction systems with equilibria is illustrated via the simulated example of the oxidation of benzyl alcohol by hypochlorite in a batch reactor [6]. Two cases are considered, namely, (i) measurements of all species in both phases, and (ii) measurements of the kinetic species and equilibrium components in the phase where the instantaneous equilibria take place (phase G).

Reaction system and available measurements

The reaction system consists of an aqueous phase (labeled G) with one kinetically controlled reaction and two instantaneous equilibria ($R_g = 3$, $R_{g,k} = 1$, $R_{g,e} = 2$):



and an organic phase (labeled L) with a single kinetically controlled reaction ($R_l = R_{l,k} = 1$):



The volume of both phases is assumed to be constant and equal to 0.5 L. The mass-transfer resistance between the two phases is described by a thin film on each side of the interface, under quasi steady-state conditions. The characteristics of these two films – surface area, mass-transfer rate laws and parameters – are assumed to be the same.

In the aqueous phase, cetyltrimethylammonium bromide (QBr) dissociates instantaneously to form the ions Q^+ and Br^- (Eq. C.9c). The Q^+ ions react instantaneously with the hypochlorite ions (OCl^-) to form QOCl (Eq. C.9b), which transfers to the organic phase. In the organic phase, benzyl alcohol ($\text{C}_6\text{H}_5\text{CH}_2\text{OH}$) in excess reacts with QOCl coming from the aqueous phase and forms benzyl aldehyde ($\text{C}_6\text{H}_5\text{CHO}$), QCl and water. The species QCl formed in the organic phase transfers to the aqueous phase, where its dissociation in ions Q^+ and Cl^- (Eq. C.9a) is kinetically observable.

The reaction and mass-transfer rate expressions as well as the equilibrium constants used for simulating the reaction schemes (C.9) and (C.10) are:

$$r_g(t) = k_{g,1} c_{\text{QCl},g}(t) - k_{g,2} c_{\text{Q}^+,g}(t) c_{\text{Cl}^-,g}(t) \quad (\text{C.11})$$

$$r_l(t) = k_l c_{\text{QOCl},l}(t) \quad (\text{C.12})$$

$$\zeta_{\text{QOCl}}(t) = k_{m,\text{QOCl}} A_l V_l (c_{\text{QOCl},g}(t) - c_{\text{QOCl},l}(t)) \quad (\text{C.13})$$

$$\zeta_{\text{QCl}}(t) = k_{m,\text{QCl}} A_l V_l (c_{\text{QCl},g}(t) - c_{\text{QCl},l}(t)) \quad (\text{C.14})$$

$$K_{g,2} = \frac{c_{\text{QOCl},g}(t)}{c_{\text{Q}^+,g}(t) c_{\text{OCl}^-,g}(t)} \quad (\text{C.15})$$

$$K_{g,3} = \frac{c_{\text{QBr},g}(t)}{c_{\text{Q}^+,g}(t) c_{\text{Br}^-,g}(t)} \quad (\text{C.16})$$

with A_l the specific interfacial area. The various sets of species present in this reaction system are listed in Table C.1. Note that $\mathcal{S}_{l,k} = \mathcal{S}_l$ with $\mathcal{S}_{l,e}$ and $\mathcal{S}_{l,c}$ empty. The $S_{g,e} =$

Table C.1 Sets of species and components involved in the oxidation of benzyl alcohol with hypochlorite.

Set	Species	Dimension
\mathcal{S}_l	$\{\text{QOCl}, \text{C}_6\text{H}_5\text{CH}_2\text{OH}, \text{QCl}, \text{C}_6\text{H}_5\text{CHO}, \text{H}_2\text{O}\}$	$S_l = 5$
$\mathcal{S}_{l,k}$	$\{\text{QOCl}, \text{C}_6\text{H}_5\text{CH}_2\text{OH}, \text{QCl}, \text{C}_6\text{H}_5\text{CHO}, \text{H}_2\text{O}\}$	$S_{l,k} = 5$
$\mathcal{S}_{l,e}$	$\{\emptyset\}$	$S_{l,e} = 0$
$\mathcal{S}_{l,c}$	$\{\emptyset\}$	$S_{l,c} = 0$
\mathcal{S}_g	$\{\text{Cl}^-, \text{QCl}, \text{Q}^+, \text{OCl}^-, \text{Br}^-, \text{QOCl}, \text{QBr}\}$	$S_g = 7$
$\mathcal{S}_{g,k}$	$\{\text{Cl}^-, \text{QCl}\}$	$S_{g,k} = 2$
$\mathcal{S}_{g,e}$	$\{\text{Q}^+, \text{OCl}^-, \text{Br}^-, \text{QOCl}, \text{QBr}\}$	$S_{g,e} = 5$
$\mathcal{S}_{g,c}$	$\{\text{Q}_c, \text{OCl}_c, \text{Br}_c\}$	$S_{g,c} = 3$
\mathcal{S}_m	$\{\text{QOCl}, \text{QCl}\}$	$S_m = 2$

5 equilibrium species in the aqueous phase G can be reduced to $S_{g,c} = 3$ equilibrium components that represent the following conserved quantities:

$$c_{\text{Q}_c,g} := c_{\text{Q}^+,g} + c_{\text{QOCl},g} + c_{\text{QBr},g} = c_{\text{Q}^+,g} (1 + K_{g,2} c_{\text{OCl}^-,g} + K_{g,3} c_{\text{Br}^-,g}) \quad (\text{C.17a})$$

$$c_{\text{OCl}_c,g} := c_{\text{OCl}^-,g} + c_{\text{QOCl},g} = c_{\text{OCl}^-,g} (1 + K_{g,2} c_{\text{Q}^+,g}) \quad (\text{C.17b})$$

$$c_{\text{Br}_c,g} := c_{\text{Br}^-,g} + c_{\text{QBr},g} = c_{\text{Br}^-,g} (1 + K_{g,3} c_{\text{Q}^+,g}), \quad (\text{C.17c})$$

where the concentrations of the equilibrium products $c_{\text{QOCl},g}$ and $c_{\text{QBr},g}$ have been replaced by their expression obtained from the thermodynamic equilibrium (Eqs. C.15 and C.16). The values of the model parameters used in the simulation are given in Table C.2.

The aqueous phase G is loaded with 0.125 kmol of OCl^- in ionic form, 0.04 kmol of QBr and 0.005 kmol of QCl. The initial amount of benzyl alcohol in the organic phase L is 0.968 kmol. The numbers of moles in both phases are simulated for 10 min and all the measured numbers of moles are corrupted with 2% zero-mean Gaussian noise with respect to the maximal number of moles of each species.

Table C.2 Kinetic, thermodynamic and surface parameters used in the simulation.

Parameter	Value	Unit	Parameter	Value	Unit
$k_{g,1}$	1.663×10^{-2}	s^{-1}	k_l	22.7	s^{-1}
$k_{g,2}$	2.5	$\text{m}^3 \text{ kmol}^{-1} \text{ s}^{-1}$	$k_{m,QOCl}$	8.02×10^{-5}	m s^{-1}
$K_{g,2}$	1.157×10^3	$\text{m}^3 \text{ kmol}^{-1}$	$k_{m,QCl}$	8.91×10^{-5}	m s^{-1}
$K_{g,3}$	0.235×10^3	$\text{m}^3 \text{ kmol}^{-1}$	A_l	200	m^{-1}

For the computation of extents, two situations with the following measurements are considered: (Case 1) all species in both phases, (Case 2) all species in phase L and the kinetic and equilibrium species in phase G.

Case 1: Measurement of all species

The extents are obtained from the measured numbers of moles using the transformations \mathcal{T}_l and \mathcal{T}_g of Section 3.2.1. For phase L, $R_l = 1$ extent of reaction and $p_m = 2$ extents of mass transfers are extracted from the $S_l = 5$ measured numbers of moles. For phase G, $R_{g,k} = 1$ extent of (kinetically controlled) reaction, $R_{g,e} = 2$ extents of equilibria, and $p_m = 2$ extents of mass transfers are extracted from the $S_g = 7$ measured numbers of moles. The measured numbers of moles as well as the extents of reaction and mass transfer are shown in Fig. C.1. Note that, since $\text{C}_6\text{H}_5\text{CH}_2\text{OH}$ is in large excess, its concentration in the organic phase is constant and thus not shown. The amount of Q^+ in the aqueous phase is also too small to be represented in Fig. C.1. In addition, since the reaction system has no outlet (batch conditions), the extents of mass transfer $\mathbf{x}_{m,g}$ and $\mathbf{x}_{m,l}$ are equal and opposite.

Case 2: Measurement of all species in phase L and of kinetic species and equilibrium components in phase G

The extents are obtained using the transformation \mathcal{T}_l of Section 3.2.1 for the organic phase L and the transformation \mathcal{T}_g of Section C.2.1 for the aqueous phase G. For phase L, $R_l = 1$ extent of reaction and $p_m = 2$ extents of mass transfers are extracted from the $S_l = 5$ measured numbers of moles. For phase G, $R_{g,k} = 1$ extent of (kinetically controlled) reaction and $p_m = 2$ extents of mass transfers are computed from the $S_{g,k} + S_{g,c} = 5$ measured numbers of moles.

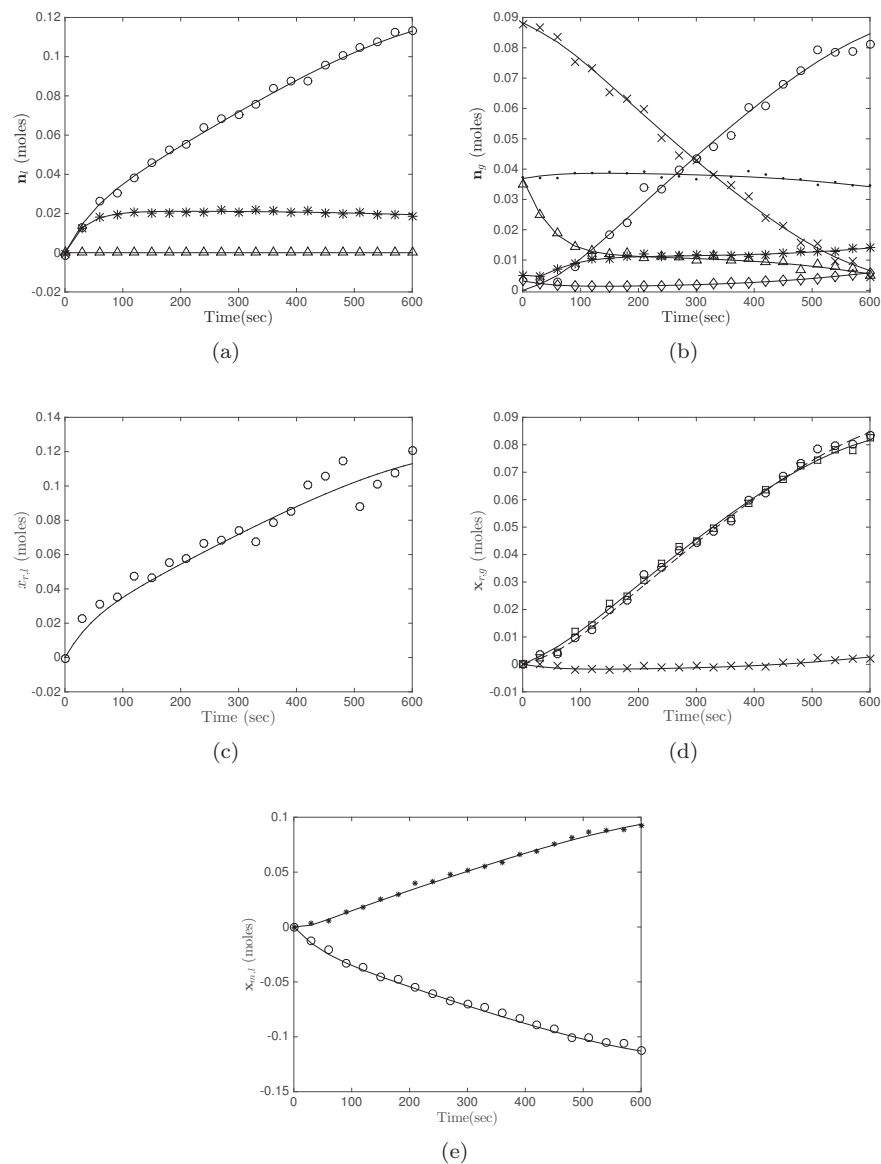


Figure C.1 Case 1: Measurement of all species. (Top (a) and (b)) simulated (noise-free, continuous lines) and measured (noisy, 2%) numbers of moles of species H_2O (o), QCl (★) and QOCl (Δ) in the organic phase and of species Cl^- (o), QCl (★), OCl^- (×), Br^- (●), QOCl (Δ) and QBr (◊) in the aqueous phase; (center (c) and (d)) Experimental (computed from measurements) and modeled (continuous and dashed lines) extents of reaction in the organic phase, with $x_{r,l}$ (o), and in the aqueous phase, with $x_{r,g,k}$ (×), $x_{r,g,e,1}$ (◻) and $x_{r,g,e,2}$ (o); (bottom (e)) Experimental and modeled (continuous lines) extents of mass transfer of QOCl (★) and QCl (o) in the organic phase.

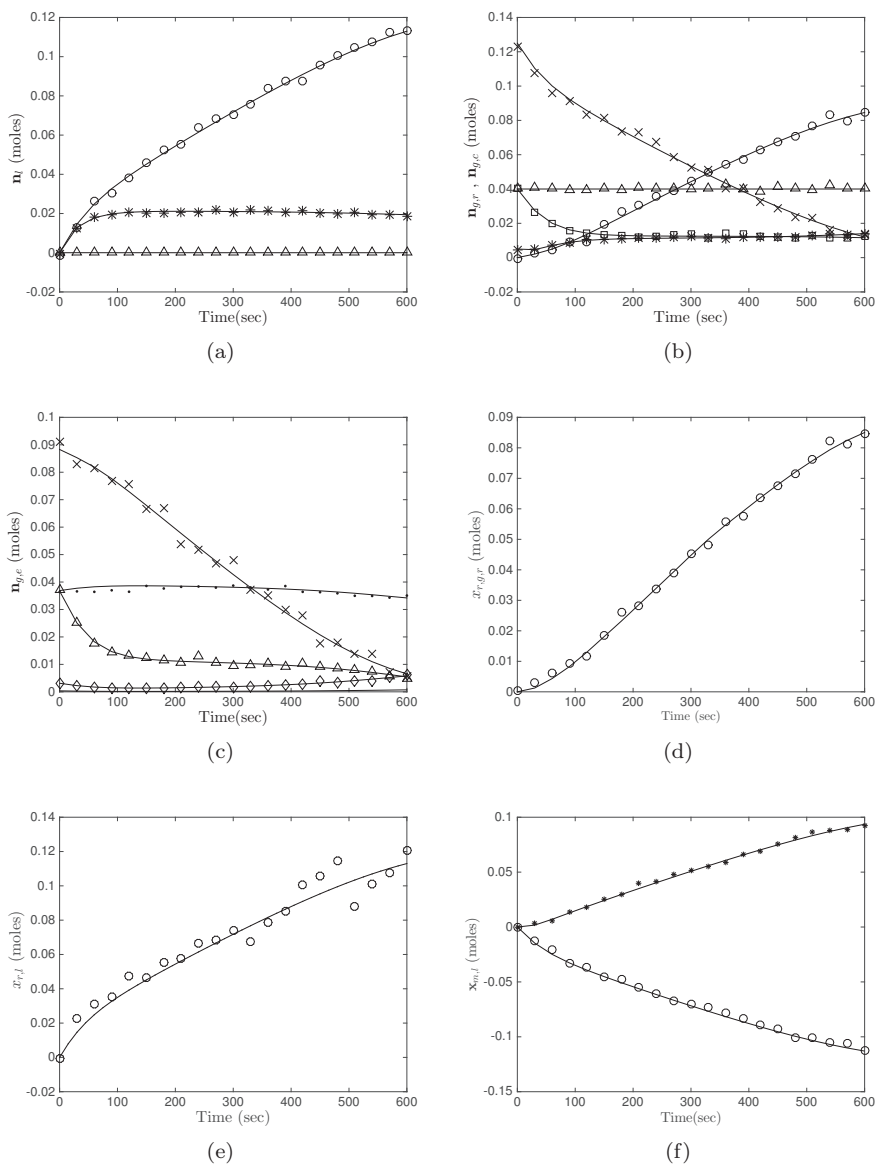


Figure C.2 Case 2: Measurement of all species in phase L and of kinetic species and equilibrium components in phase G. (Top (a) and (b)) Simulated (noise-free, continuous lines) and measured (noisy, 2%) numbers of moles of species H_2O (o), QCl (\star) and $QOCl$ (Δ) in the organic phase, and of kinetic species Cl^- (o) and QCl (\star) and equilibrium components Q_c (\square), OCl_c (\times) and Br_c (Δ) in the aqueous phase; (center (c)) Simulated (continuous lines) and reconstructed numbers of moles of the equilibrium species OCl^- (\times), Br^- (\bullet), $QOCl$ (Δ) and QBr (\diamond) in the aqueous phase; (center (d) and bottom (e)) Experimental (computed from measurements) and modeled (continuous lines) extents of (kinetically controlled) reaction in the aqueous and organic phases; (bottom (f)) Experimental and modeled (continuous lines) extents of mass transfer of species $QOCl$ (\star) and QCl (o) in the organic phase.

C.4 Summary

The presence of fast reactions (instantaneous equilibria) affects the other rate processes such as kinetically controlled reactions and mass transfers. This chapter has introduced a procedure for decoupling the chemical reaction system with instantaneous equilibria using the measurements of only the kinetic species and equilibrium components. This allows the computation of the extents of only the kinetically controlled reactions, that can then be modeled independently of the instantaneous equilibria. A detailed procedure for extent-based approach in the presence of instantaneous equilibria has been documented by Srinivasan et al. [71].

Appendix D

Extent-based Calibration of Spectroscopic Data

The recent development of Process Analytical Technologies (PAT) has opened up new avenues for exploiting spectroscopic data, which indirectly provide non-destructive concentration measurements of the species in-situ/online during the course of a reaction [80]. Accurate spectroscopic measurements are available at low cost and short sampling times, and are free of any delay. In addition, most spectroscopic measurements can be considered linear with respect to the concentrations of the absorbing species. The technology of spectrometers and fiber-optic probes has improved significantly in recent years. Furthermore, numerical methods facilitate the analysis of large sets of multivariate data. For these reasons, online spectroscopic measurements are nowadays widely used, probably more so than direct concentration measurements [65, 25].

In absorption spectroscopy, calibration models based on Principal Component Regression (PCR) [47] or Partial Least Squares (PLS) [37] are used to predict the concentrations of absorbing species from an absorbance spectrum measured in a sample [31, 46]. Calibration involves two steps. In the calibration step, the calibration model is built using pairs of measured concentrations and absorbance spectra. This construction requires defining a certain number of latent variables that represent abstract building blocks. In the prediction step, the calibration model is applied to the absorbance spectrum measured in a new sample to predict the concentrations of the calibrated species.

Since calibration models require that the new spectrum lies in the space spanned by the calibration set (space-inclusion condition), it is important to properly design the calibration set [3]. This condition translates the fact that the experimental conditions used for the calibration set should mimic as much as possible the experimental conditions of the prediction experiment, *inter alia*, in terms of concentration and temperature ranges [38, 49]. One approach consists in preparing calibration samples of various independent compositions and measuring the absorbance spectrum of each sample. The resulting non-reacting calibration data are sometimes qualified as ‘static’ in the sense that the composition of each calibration sample does not vary with time [79]. This way of constructing the calibration set has the drawback that it requires numerous calibration mixtures and that none of the calibration samples contains the short-lived absorbing intermediates that are often found in real samples taken during the course of a reaction. This can cause a more or less severe violation of the space-inclusion condition. An alternative approach consists in taking calibration samples during the course of a preliminary experiment, based on the chemical process at hand, and measuring pairs of

concentrations and absorbance spectra for each of these samples. With such a design, the intermediate species that are produced by the reactions are included, and the need to vary them externally disappears [3]. Such a ‘dynamic’ calibration system is particularly suitable for industry, where reference samples taken during the course of the reaction under normal operating conditions are usually available and can be used for calibration.

When reacting calibration data are used, the calibration model can be constructed on two types of calibration pairs. A first possibility consists in taking measured concentrations and absorbance spectra as calibration pairs; in such a case, the calibration model predicts concentrations from a new spectrum. Alternatively, the calibration model can be constructed on pairs of concentration and spectroscopic measurements containing only the reaction and mass-transfer contributions, that is, the contribution of the inlet and outlet flows is removed from the measurements. The calculation of these pairs requires the pretreatment of concentrations and spectroscopic data in so-called *vessel reaction- and mass-transfer-variant (vRMV) form*.

In this Appendix, we show the procedure for calibrating and then predicting the concentrations using the *vRMV*- and the *vRV*- forms, for a two-phase reaction system (G-L) with reactions only in phase L and steady-state mass transfer between phases.

D.1 Factorization of Spectroscopic Data from Reaction Systems

The starting point of this quantitative analysis of spectroscopic data is Beer’s law. Assume that an absorbance signal is measured in phase L by a spectrometer, which produces data that have a linear response with respect to concentrations. Let $\mathbf{a}(t)$ be a W -dimensional absorbance spectrum measured for a unit pathlength at W wavelengths and at time t . Beer’s law allows writing:

$$\mathbf{a}(t) = \mathbf{E}^T \mathbf{c}_{l,a}(t) \quad (\text{D.1})$$

where $\mathbf{c}_{l,a}(t)$ is the S_a -dimensional vector of concentrations and \mathbf{E} is the $S_a \times W$ matrix containing the pure component spectra (molar absorptivities) of the S_a absorbing species ($S_a \leq S_l$). The subscript $(\cdot)_a$ indicates a quantity associated with the absorbing species.

D.1.1 Standard factorization

The reconstruction of the numbers of moles $\mathbf{n}_l(t)$ from the extents $\mathbf{x}_r(t)$, $\mathbf{x}_{m,l}(t)$, $\mathbf{x}_{in,l}(t)$ and $x_{ic,l}(t)$ given in Eq. (3.40) can be used to compute the concentrations $\mathbf{c}_{l,a}(t)$ of the S_a absorbing species at time t :

$$\mathbf{c}_{l,a}(t) = \frac{1}{V_l(t)} \left(\mathbf{N}_a^T \mathbf{x}_r(t) + \mathbf{W}_{m,l,a} \mathbf{x}_{m,l}(t) + \mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,a} x_{ic,l}(t) \right) \quad (\text{D.2})$$

Combining Eqs. (D.1) and (D.2), together with $\mathbf{n}_{l0,a} = \mathbf{c}_{l0,a} V_{l0}$, leads to the so-called *factorization of spectroscopic data* [2, 3] given by:

$$\mathbf{a}(t) = \frac{1}{V_l(t)} \left(\mathbf{E}^T \mathbf{N}_a^T \mathbf{x}_r(t) + \mathbf{A}_m^T \mathbf{x}_{m,l}(t) + \mathbf{A}_{in}^T \mathbf{x}_{in,l}(t) + \mathbf{a}_0 V_{l0} x_{ic,l}(t) \right) \quad (\text{D.3})$$

with $\mathbf{A}_m^T = \mathbf{E}^T \mathbf{W}_{m,l,a}$ the $W \times p_m$ spectra of the species transferring between the two phases (mass-transfer spectra), $\mathbf{A}_{in}^T = \mathbf{E}^T \mathbf{W}_{in,l,a}$ the $W \times p_l$ spectra of the inlet flows (inlet spectra) and $\mathbf{a}_0 = \mathbf{E}^T \mathbf{c}_{l0,a}$ the W -dimensional initial spectrum.

D.1.2 Factorization in vRMV-form

Similarly to Eq. (3.42), if the inlet and outlet flow rates $\mathbf{u}_{in,l}(t)$ and $u_{out,l}(t)$ and the mass $m_l(t)$ of the reaction mixture are known, the extents $\mathbf{x}_{in,l}(t)$ and the variable $x_{ic}(t)$ can be calculated by integration of the differential equations (3.39c) and (3.39d). It follows that, if $V_l(t)$, $\mathbf{W}_{in,l,a}$ and $\mathbf{n}_{l0,a}$ are also known, Eq. (D.2) can be written in RMV-form as:

$$\mathbf{c}_{l,a}^{vRMV}(t) := \mathbf{c}_{l,a}(t) - \frac{1}{V_l(t)} \left(\mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,a} x_{ic,l}(t) \right) = \frac{1}{V_l(t)} \begin{bmatrix} \mathbf{N}_a^T & \mathbf{W}_{m,l,a} \end{bmatrix} \begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{m,l}(t) \end{bmatrix} \quad (\text{D.4})$$

In addition, if the initial spectrum \mathbf{a}_0 and the inlet spectra \mathbf{A}_{in} are known, Eq. (D.3) can be rearranged in vRMV-form that accounts only for the spectroscopic contributions of the reactions and mass transfers:

$$\mathbf{a}^{vRMV}(t) := \mathbf{a}(t) - \frac{1}{V_l(t)} \left(\mathbf{A}_{in}^T \mathbf{x}_{in,l}(t) + \mathbf{a}_0 V_{l0} x_{ic,l}(t) \right) = \frac{1}{V_l(t)} \begin{bmatrix} \mathbf{E}^T \mathbf{N}_a^T & \mathbf{A}_m^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{m,l}(t) \end{bmatrix} \quad (\text{D.5})$$

D.1.3 Factorization in RV-form

When the p_m extents of mass transfer are calculated from the numbers of moles measured in gas phase, the concentration and spectroscopic contributions of the mass transfers can be subtracted from Eqs. (D.4) and (D.5), which leads to the following equations:

$$\mathbf{c}_{l,a}^{vRV}(t) := \mathbf{c}_{l,a}(t) - \frac{1}{V_l(t)} \left(\mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,a} x_{ic,l}(t) + \mathbf{W}_{m,l,a} \mathbf{x}_{m,l}(t) \right) = \frac{1}{V_l(t)} \mathbf{N}_a^T \mathbf{x}_r(t) \quad (\text{D.6})$$

$$\mathbf{a}^{vRV}(t) := \mathbf{a}(t) - \frac{1}{V_l(t)} \left(\mathbf{A}_{in}^\top \mathbf{x}_{in,l}(t) + \mathbf{a}_0 V_{l0} x_{ic,l}(t) + \mathbf{A}_m^\top \mathbf{x}_{m,l}(t) \right) = \frac{1}{V_l(t)} \mathbf{E}^\top \mathbf{N}_a^\top \mathbf{x}_r(t) \quad (\text{D.7})$$

Eqs. (D.6) and (D.7) represent the RV-forms of Eqs. (D.2) and (D.3), respectively.

For homogeneous reaction systems, there is no mass-transfer contribution and Eq. (D.6) simplifies accordingly, namely, $\mathbf{W}_{m,l,a} \mathbf{x}_{m,l}(t) = \mathbf{0}_{S_a}$. It follows that the equation to compute the spectroscopic contribution of the reactions $\mathbf{a}^{vRV}(t)$ is identical to the expression that holds when all transferring species are spectroscopically silent (non-absorbing), namely, when $\mathbf{A}_m^\top = \mathbf{0}_{W \times p_m}$.

D.2 Prediction of Calibrated Concentrations from Spectroscopic Data

Concentrations in reaction systems can be predicted from measured absorbance spectra using multivariate calibration. The standard approach for building a calibration model consists in taking measured concentrations and absorbance spectra as calibration pairs. Other choices are possible and result in different calibration models.

D.2.1 Standard calibration

The calibration proceeds in two steps. In the calibration step, a calibration model is constructed using \mathbf{f} latent variables. The calibration set consists of H pairs of vectors that include W -dimensional absorbance spectra $\mathbf{a}_{cal}(t)$ and S_c -dimensional concentrations $\mathbf{c}_{l,cal}(t)$, where t is one of the H sampling times and S_c the number of species available for calibration ($S_c \leq S_l$). The subscript $(\cdot)_{cal}$ denotes a quantity associated with the calibration experiment and the subscript $(\cdot)_c$ indicates that the vector or matrix original dimension comprised of S_l species is reduced to S_c calibrated species. A calibration model can be expressed by the calibration matrix \mathbf{F} of dimensions $S_c \times W$ comprising S_c rows of *prognostic* vectors of dimension W that describe the relation between all H pairs $(\mathbf{c}_{l,cal}(t), \mathbf{a}_{cal}(t))$. This is illustrated in the central part of Fig. D.1.

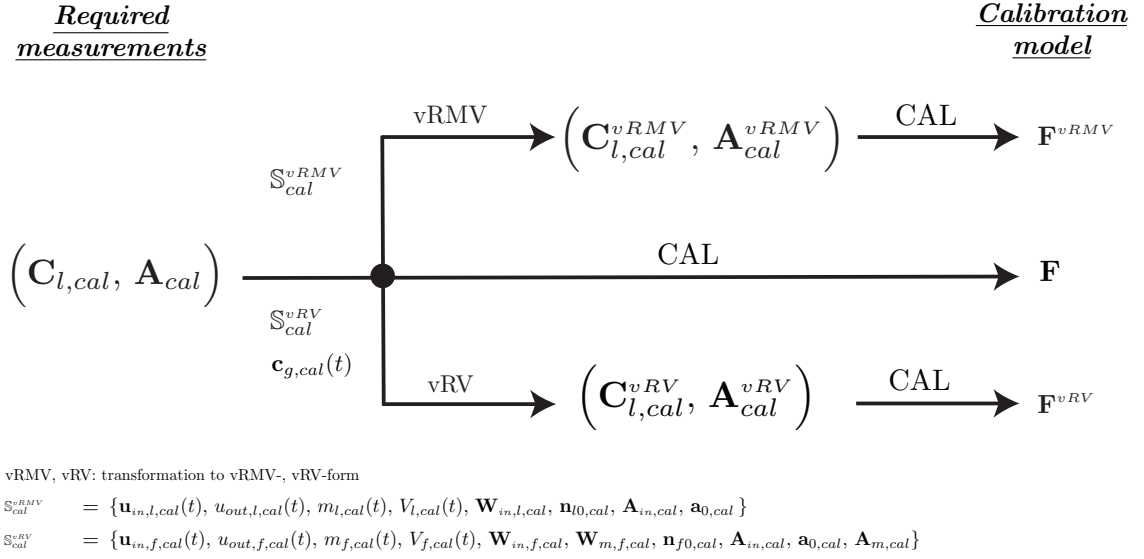


Figure D.1 Construction of the calibration model for standard calibration, calibration in RMV-form and calibration in RV-form.

The method for constructing the matrix \mathbf{F} depends on the calibration technique used. All calibration techniques require turning the time dependence t of the vectors $\mathbf{a}_{cal}(t)$ and $\mathbf{c}_{l,cal}(t)$ into an additional dimension. Hence, the H pairs of vectors $\mathbf{a}_{cal}(t)$ and $\mathbf{c}_{l,cal}(t)$ are transposed and stacked vertically in time-resolved matrices \mathbf{A}_{cal} ($H \times W$) and $\mathbf{C}_{l,cal}$ ($H \times S_c$), respectively. With PCR, the matrix \mathbf{A}_{cal} is first approximated as $\bar{\mathbf{U}} \bar{\mathbf{S}} \bar{\mathbf{V}}^T$, where $\bar{\mathbf{U}}$ ($H \times \mathbf{f}$) and $\bar{\mathbf{V}}$ ($W \times \mathbf{f}$) contain \mathbf{f} principal component vectors, $\bar{\mathbf{S}}$ ($\mathbf{f} \times \mathbf{f}$) is a \mathbf{f} -dimensional diagonal matrix of principal component factors, and \mathbf{f} is the number of principal components (also called latent variables). The calibration matrix \mathbf{F} is calculated as $\mathbf{C}_{l,cal}^T \bar{\mathbf{U}} \bar{\mathbf{S}}^{-1} \bar{\mathbf{V}}^T$. With PLS, the calculation of the calibration matrix \mathbf{F} is more complex and calls for an iterative scheme [45].

In the prediction step, the calibration model is used to predict the S_c calibrated concentrations $\hat{\mathbf{c}}_{l,c}(t)$ from the W -dimensional absorbance spectrum $\mathbf{a}(t)$ measured at time t of a prediction (or test) experiment:

$$\hat{\mathbf{c}}_{l,c}(t) = \mathbf{F} \mathbf{a}(t) \quad (\text{D.8})$$

D.2.2 Calibration in vRMV-form

When calibration models are constructed using only the spectral contributions of the reactions and mass transfers, the pairs of calibration vectors $(\mathbf{c}_{l,cal}^{vRMV}(t), \mathbf{a}_{cal}^{vRMV}(t))$ are given by Eqs. (D.4) and (D.5). Note that the pretreatment of Eqs. (D.4) and (D.5) requires the knowledge of the inlets $(\mathbf{u}_{in,l,cal}(t), \mathbf{W}_{in,l,cal}, \mathbf{A}_{in,cal})$, the outlet $(u_{out,l,cal}(t))$

and the initial conditions $(\mathbf{n}_{l0,cal}, \mathbf{a}_{0,cal})$, as well as the mass and volume $(m_{l,cal}(t), V_{l,cal}(t))$ of the calibration mixture as illustrated in the upper part of Fig. D.1. The construction of the calibration matrix \mathbf{F}^{vRMV} is similar to the standard calibration, with the matrices \mathbf{A}_{cal} and $\mathbf{C}_{l,cal}$ replaced by \mathbf{A}_{cal}^{vRMV} ($H \times W$) and $\mathbf{C}_{l,cal}^{vRMV}$ ($H \times S_c$), respectively.

To predict the concentrations of the calibrated species, the spectroscopic contributions of the reactions and mass transfers are computed using Eq. (D.5) from the absorbance spectrum $\mathbf{a}(t)$ measured for a prediction experiment. The resulting vector $\mathbf{a}^{vRMV}(t)$ is used by the calibration model to predict the S_c concentrations in RMV-form, $\hat{\mathbf{c}}_{l,c}^{vRMV}(t)$, for all time instants of the prediction experiment.

D.2.3 Calibration in vRV-form

If gas-phase measurements can be used to compute the extents of mass transfer in the liquid phase, the calibration model can be built on the reaction contribution only by removing the contribution due to mass transfer. The pairs of calibration vectors $(\mathbf{c}_{l,cal}^{vRV}(t), \mathbf{a}_{cal}^{vRV}(t))$ are obtained by pretreatment using Eqs. (D.6) and (D.7), which requires the knowledge of the experimental conditions at all calibration times (lower part of Fig. D.1). The calibration matrix \mathbf{F}^{vRV} is constructed in a similar way as for standard calibration, with the matrices \mathbf{A}_{cal} and $\mathbf{C}_{l,cal}$ replaced by \mathbf{A}_{cal}^{vRV} ($H \times W$) and $\mathbf{C}_{l,cal}^{vRV}$ ($H \times S_c$), respectively.

In the prediction step, the absorbance spectrum $\mathbf{a}(t)$ measured during a prediction experiment is transformed to the RV-form $\mathbf{a}^{vRV}(t)$ using Eq. (D.7). This spectroscopic contribution is used to predict the S_c concentrations in RV-form, $\hat{\mathbf{c}}_{l,c}^{vRV}(t)$. Note that this procedure is also used when there are no transferring species in the calibration set.

For homogeneous reaction systems, standard calibration is based on the pairs $(\mathbf{c}_{cal}(t), \mathbf{a}_{cal}(t))$. For calibration in RV-form, $\mathbf{c}_{cal}^{vRV}(t)$ is obtained using Eq. (D.6) without the term $\mathbf{W}_{m,l}\mathbf{x}_{m,l}(t)$, while $\mathbf{a}_{cal}^{vRV}(t)$ results from Eq. (D.7) without the term $\mathbf{A}_m^T \mathbf{x}_{m,l}(t)$. In the prediction step, the pretreated vector $\mathbf{a}^{vRV}(t)$ is then used to predict the concentrations $\hat{\mathbf{c}}_c^{vRV}(t)$ associated with the reactions.

D.2.4 Choice of calibration model

Which path to choose in Fig. D.1 depends on whether or not the spectrum $\mathbf{a}(t)$ of the prediction experiment lies in the subspace spanned by the calibration set. This space-inclusion condition has been studied in [3, 2]. For Path 1, the space-inclusion condition $\mathbf{a}(t) \in \text{Im}(\mathbf{A}_{cal})$ can be verified by checking that the Euclidean norm $\epsilon(\mathbf{a}(t), \mathbf{A}_{cal})$ of the projection error of $\mathbf{a}(t)$ on the row space of \mathbf{A}_{cal} is nearly zero. Such a norm can be computed as $\epsilon(\mathbf{a}(t), \mathbf{A}_{cal}) = \|\mathbf{a}^T(t)(\mathbf{I}_W - \mathbf{A}_{cal}^+ \mathbf{A}_{cal})\|_2$. If this condition is not satisfied,

Path 2 can be followed, which implies that the calibration and linear transformation steps are performed in RMV-form. For Path 2, the space-inclusion condition becomes $\mathbf{a}^{vRMV}(t) \in \text{Im}(\mathbf{A}_{cal}^{vRMV})$ and can be verified by calculating $\epsilon(\mathbf{a}^{vRMV}(t), \mathbf{A}_{cal}^{vRMV})$. If the space-inclusion condition is still not fulfilled, additional independent measurements, such as gas-phase measurements, are required to take Path 3 and use a calibration model in RV-form. For such a calibration, the space-inclusion condition is $\mathbf{a}^{vRV}(t) \in \text{Im}(\mathbf{A}_{cal}^{vRV})$ and can be checked by computing $\epsilon(\mathbf{a}^{vRV}(t), \mathbf{A}_{cal}^{vRV})$. If none of these space-inclusion conditions can be satisfied, the calibration set must be re-designed and it might be necessary to choose another set of S_c calibrated species.

D.3 Simulated Example

In this section, we illustrate the extent-based calibration technique via simulation of the acetoacetylation of pyrrole system introduced earlier in Chapter 3.

Simulated measurements

All simulated experiments are conducted in an isothermal continuous stirred-tank reactor (CSTR) at 50 °C. The reactor is initially loaded with $V_0 = 1$ L of a solution of pyrrole in toluene and pure diketene is fed ($p = 1$) at a constant mass flow rate with the inlet composition $\mathbf{w}_{in} = [0 \ M_{w,B}^{-1} \check{w}_{in,B} \ 0 \ 0 \ 0 \ 0 \ 0]^T$, where $M_{w,B} = 84.08$ g mol⁻¹ and $\check{w}_{in,B} = 1$. Since the density and the volume of the reaction mixture are assumed to be constant throughout the course of the reaction, $u_{out}(t) = u_{in}(t)$.

Each concentration is corrupted by additive zero-mean normally distributed noise with standard deviation corresponding to 1 % of the maximal concentration. Absorbance spectra are assumed to be measured in the mid-IR region (500-1500 cm⁻¹) with 4 cm⁻¹ resolution, thus leading to $W = 250$ wavenumbers monitored simultaneously. The spectrum of the solvent (toluene) is assumed to be treated as background spectrum. All species involved in the reactions are absorbing, so that $S_a = S$ and $\mathbf{c}_a(t) = \mathbf{c}(t)$. The absorbance spectra $\mathbf{a}(t)$ are generated using Eq. (D.1) from the simulated concentrations (corrupted with 1% noise) and the (7×250) -dimensional pure component spectra \mathbf{E} chosen to resemble the true ones. The pure component spectra are shown in Fig. D.2.

Since the quantities $u_{in}(t)$, $u_{out}(t)$, $m(t)$ and $V(t)$ are known, the extent of inlet flow $x_{in}(t)$ and initial condition x_{ic} can be computed. Under the condition that all experimental conditions (including the inlet spectrum of diketene) are known, the concentration and spectral contributions associated with the inlet and outlet can be removed from the vectors $\mathbf{c}(t)$ and $\mathbf{a}(t)$ to generate $\mathbf{c}^{vRV}(t)$ and $\mathbf{a}^{vRV}(t)$ using Eqs. (D.6) and (D.7).

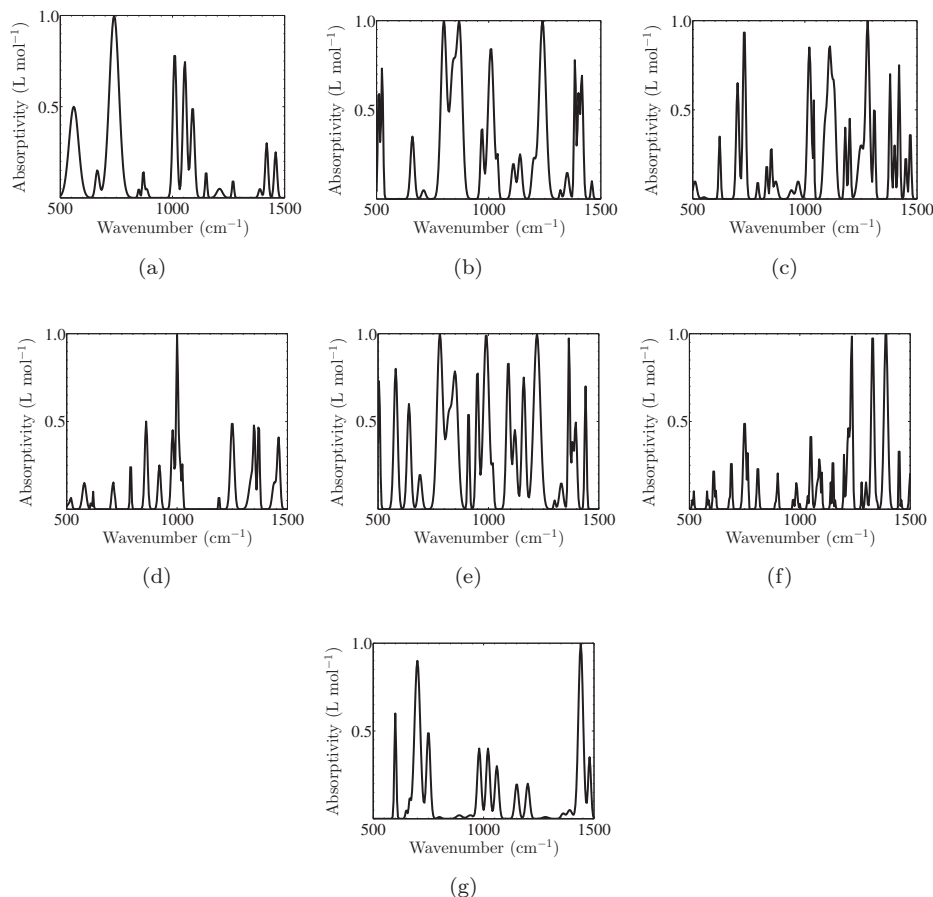


Figure D.2 Acetoacetylation of pyrrole: Simulated mid-IR (500-1500 cm^{-1}) pure component spectra (absorptivities) of (a) pyrrole, (b) diketene, (c) 2-acetoacetyl pyrrole, (d) dehydroacetic acid, (e) oligomer, (f) by-product and (g) pyridine, normalized to a unit peak height.

Calibration experiment

In the calibration experiment, the reactor is initially loaded with a solution of composition $\mathbf{c}_{0,cal} = [0.92 \ 0.07 \ 0.10 \ 0.02 \ 0 \ 0 \ 4.50]^T \text{ mol L}^{-1}$ of mass 919.98 g. Pure diketene is fed at a constant mass flow rate of 3.65 g min^{-1} , and the reactor content is continuously withdrawn at the same flow rate. The concentrations and absorbance spectra are generated according to the procedure described in Section D.3.

The calibration set is built by taking $H = 15$ absorbance spectra at the equally spaced times $t = 0, 10, \dots, 140 \text{ min}$, that is, with a sampling time of 10 min. In parallel, it is assumed that samples are taken from the reaction mixture and concentrations of all species except the catalyst are determined by an offline measurement technique. This sampling procedure results in 15 pairs of calibration vectors $(\mathbf{c}_{cal}(t), \mathbf{a}_{cal}(t))$ for the S_c

= 6 species {A, B, C, D, E, F}. As all experimental conditions (including the inlet spectrum) are assumed to be known, the vectors $\mathbf{c}_{cal}^{vRV}(t)$ and $\mathbf{a}_{cal}^{vRV}(t)$ are also available.

The (6×250) -dimensional calibration matrices \mathbf{F} and \mathbf{F}^{vRV} are constructed by PCR using the pairs of calibration vectors $(\mathbf{c}_{cal}(t), \mathbf{a}_{cal}(t))$ and $(\mathbf{c}_{cal}^{vRV}(t), \mathbf{a}_{cal}^{vRV}(t))$, respectively. One extra latent variable is used to account for the effect of noise, that is, $\mathbf{f} = 7$ latent variables instead of $S_c = 6$ for \mathbf{F} , and $\mathbf{f} = 5$ latent variables instead of $R = 4$ for \mathbf{F}^{vRV} .

Prediction experiment

In the prediction experiment, the reactor is initially loaded with a solution of composition $\mathbf{c}_0 = [0.72 \ 0.09 \ 0.10 \ 0.02 \ 0 \ 0 \ 5.00]^T \text{ mol L}^{-1}$ of mass 923.46 g. Pure diketene is fed at a constant mass flow rate of 4.40 g min^{-1} , and the reactor content is continuously withdrawn at the same flow rate. The absorbance spectra are generated according to the procedure described in Section D.3.

The prediction set is built by taking absorbance spectra every 1 min for a duration of 145 min. The resulting measured absorbance spectra with 1.5% noise are shown in Fig. D.3. As all experimental conditions (including the inlet spectrum) are assumed to be known, the vector $\mathbf{a}^{vRV}(t)$ is also available. Fig. D.4 shows the simulated and predicted concentrations using \mathbf{F}^{vRV} .

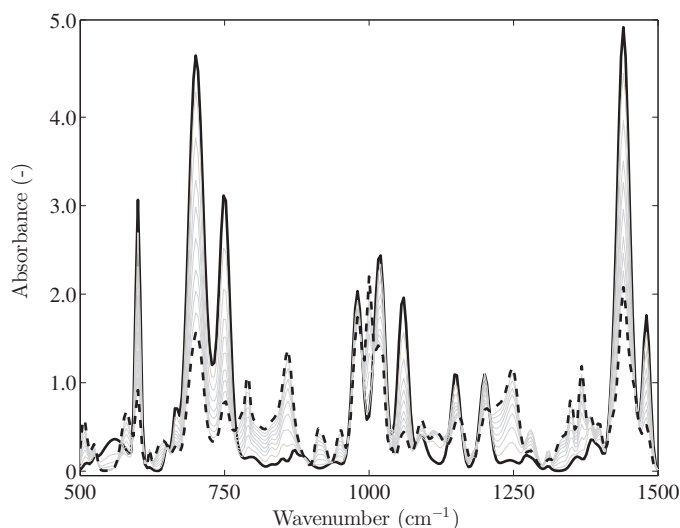


Figure D.3 Acetoacetylation of pyrrole: Mid-IR absorbance spectra $\mathbf{a}(t)$ with 1.5% noise measured every minute but plotted every 15 min. The black continuous and dashed lines indicate the first spectrum and the last spectrum, respectively.

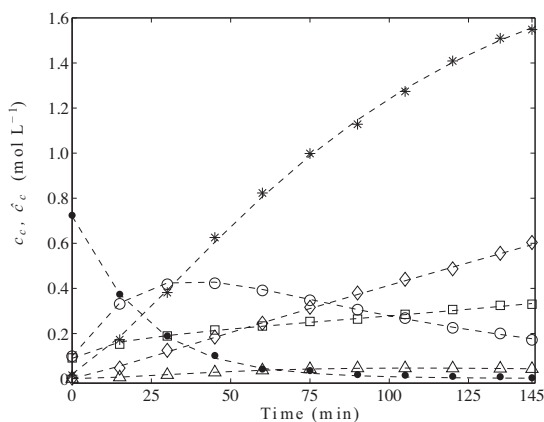


Figure D.4 Prediction experiment for acetoacetylation of pyrrole: Simulated concentrations (dashed line) and concentrations predicted via \mathbf{F}^{vRV} (PCR model) for species A (●), B (□), C (○), D (*), E (◇) and F (△);

D.4 Summary

PCR and PLS calibration techniques are widely used to predict concentrations from spectroscopic measurements. In this chapter, a procedure for building calibration models using vessel extents has been introduced. The use of extents leads to two different forms of calibration pairs, namely, the $vRMV$ -form and vRV -form. These calibration pairs requires pretreatment of calibration data and becomes a requirement for calibration and prediction when the space inclusion condition in terms of the usual concentration and absorbance data is no longer valid. A detailed procedure for building calibration models and prediction of concentrations using extents has been documented by Billeter et al. [17].

References

- [1] M. Aggarwal, S. Balaji, and B. E. Ydstie. Invariant based modeling and control of multi-phase reactor systems. *J. Process Contr.*, 21(10):1390–1406, 2011.
- [2] M. Amrhein. *Reaction and Flow Variants/Invariants for the Analysis of Chemical Reaction Data*. Doctoral thesis No. 1861, EPFL Lausanne, Switzerland, 1998.
- [3] M. Amrhein, B. Srinivasan, D. Bonvin, and M. M. Schumacher. Calibration of spectral reaction data. *Chemometrics and Intelligent Laboratory Systems.*, 46(2): 249–264, 1999.
- [4] M. Amrhein, N. Bhatt, B. Srinivasan, and D. Bonvin. Extents of reaction and flow for homogeneous reaction systems with inlet and outlet streams. *AIChE J.*, 56(11): 2873–2886, 2010.
- [5] R. Aris and R. H. S. Mah. Independence of chemical reactions. *Industrial & Engineering Chemistry Fundamentals*, 2(2):90–94, 1963.
- [6] S. Asai, H. Nakamura, and T. Sumita. Oxidation of benzyl alcohol using hyochlorite ion via phase-transfer catalysis. *AIChE J.*, 40(12):2028–2033, 1994.
- [7] O. A. Asbjørnsen. Reaction invariants in the control of continuous chemical reactors. *Chem. Eng. Sci.*, 27:709–717, 1972.
- [8] O. A. Asbjørnsen and M. Fjeld. Response modes of continuous stirred tank reactors. *Chem. Eng. Sci.*, 25:1627–1636, 1970.
- [9] M. Baldea and I. Harjunkski. Integrated production scheduling and process control: A systematic review. *Comp. Chem. Eng.*, 71:377–390, 2014.
- [10] Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, New York, 1974.

-
- [11] A. Bardow and W. Marquardt. Incremental and simultaneous identification of reaction kinetics: Methods and comparison. *Chem. Eng. Sci.*, 59(13):2673, 2004.
 - [12] G. Bastin and J. Lévine. On state accessibility in reaction systems. *IEEE Transactions Automatic Control*, 38(5):733–742, 1993.
 - [13] N. Bhatt. *Extents of Reaction and Mass Transfer in the Analysis of Chemical Reaction Systems*. PhD thesis, No. 5028, EPFL, Lausanne, Switzerland, 2011.
 - [14] N. Bhatt, M. Amrhein, and D. Bonvin. Extents of reaction, mass transfer and flow for gas-liquid reaction systems. *Ind. Eng. Chem. Res.*, 49(17):7704–7717, 2010.
 - [15] N. Bhatt, M. Amrhein, and D. Bonvin. Incremental identification of reaction and mass-transfer kinetics using the concept of extents. *Ind. Eng. Chem. Res.*, 50(23):12960–12974, 2011.
 - [16] N. Bhatt, N. Kerimoglu, M. Amrhein, W. Marquardt, and D. Bonvin. Incremental identification for reaction systems - A comparison between rate-based and extent-based approaches. *Chem. Eng. Sci.*, 83:24–38, 2012.
 - [17] J. Billeter, S. Srinivasan, and D. Bonvin. Extent-based kinetic identification using spectroscopic measurements and multivariate calibration. *Analytica Chimica Acta*, 767:21–34, 2013.
 - [18] D. Bonvin and D. W. T. Rippin. Target factor analysis for the identification of stoichiometric models. *Chem. Eng. Sci.*, 45(12):3417–3426, 1990.
 - [19] G. E. P. Box and N. R. Draper. *Empirical Model-building and Response Surfaces*. John Wiley, New York, 1987.
 - [20] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
 - [21] M. Brendel. *Incremental Identification of Complex Reaction Systems*. PhD thesis, RWTH Aachen, Germany, 2005.
 - [22] M. Brendel, A. Mhamdi, D. Bonvin, and W. Marquardt. An incremental approach for the identification of reaction kinetics. In *IFAC ADCHEM'04*, Hong-Kong, China, 2004.
 - [23] M. Brendel, D. Bonvin, and W. Marquardt. Incremental identification of kinetic models for homogeneous reaction systems. *Chem. Eng. Sci.*, 61(16):5404–5420, 2006.
 - [24] G. A. Bunin, G. François, and D. Bonvin. From discrete measurements to bounded gradient estimates: A look at some regularizing structures. *Ind. Eng. Chem. Res.*,

- 52(35):12500–12513, 2013.
- [25] W. Chew and P. Sharratt. Trends in process analytical technology. *Anal. Methods*, 2:1412–1438, 2010.
- [26] C. M. Crowe. Data reconciliation – Progress and challenges. *J. Process Contr.*, 6(2):89–98, 1996.
- [27] M. Dabros, M. Amrhein, D. Bonvin, I. W. Marison, and U. von Stockar. Data reconciliation of concentration estimates from mid-infrared and dielectric spectral measurements for improved on-line monitoring of bioprocesses. *Biotech. Progress*, 25(2):578–588, 2009.
- [28] D. Dochain and L. Chen. Local observability and controllability of stirred tank reactors. *J. Process Contr.*, 2(3):139–144, 1992.
- [29] D. Dochain, F. Couenne, and C. Jallut. Enthalpy based modelling and design of asymptotic observers for chemical reactors. *Int. J. Control*, 82(8):1389–1403, 2009.
- [30] F.T. Edgar, D. M. Himmelblau, and L. Lasdon. *Optimization of chemical processes*. McGraw-Hill Book Company, 1989.
- [31] G.M. Escandar, N.K.M. Faber, H.C. Goicoechea, A.M. de la Pena, A.C. Olivieri, and R.J. Poppi. Second- and third-order multivariate calibration: Data, algorithms and applications. *Trac-Trends Anal. Chem.*, 26:752–765, 2007.
- [32] A. Favache and D. Dochain. Thermodynamics and chemical systems stability: The CSTR case study revisited. *J. Process Contr.*, 19(3):371–379, 2009.
- [33] M. Fjeld, O. A. Asbjørnsen, and K. J. Åström. Reaction invariants and their importance in the analysis of eigenvectors, state observability and controllability of the continuous stirred tank reactor. *Chem. Eng. Sci.*, 29:1917–1926, 1974.
- [34] J. C. Friedly. Extent of reaction in open systems with multiple heterogeneous reactions. *AIChE J.*, 37(5):687–693, 1991.
- [35] J. C. Friedly. Reaction coordinates for heterogeneous flow reactors: Physical interpretation. *AIChE J.*, 42(10):2987–2989, 1996.
- [36] S. B. Gadewar, M. F. Doherty, and M. F. Malone. Reaction invariants and mole balances for plant complexes. *Ind. Eng. Chem. Res.*, 41(16):3771–3783, 2002.
- [37] P. Geladi and B.R. Kowalski. Partial least-squares regression - A tutorial. *Analytica Chimica Acta*, 185:1–17, 1986.

-
- [38] P. J. Gemperline. *Practical Guide to Chemometrics*. Taylor and Francis, Boca Raton, USA, 2006.
- [39] S. Gruner, M. Mangold, and A. Kienle. Dynamics of reaction separation processes in the limit of chemical equilibrium. *AIChE J.*, 52(3):1010–1026, 2006.
- [40] L. G. Hammarstrom. Control of chemical reactors in the subspace of reaction and control variants. *Chem. Eng. Sci.*, 34:891–899, 1979.
- [41] N. H. Hoang, D. Dochain, and B. E. Ydstie. Partial inventory control of the cstr via reaction-dependent generalized inventories. *IFAC Proceedings Volumes*, 47(3): 9123–9128, 2014.
- [42] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
- [43] C. K. Liew. Inequality constrained least-squares estimation. *J. of the American Statistical Association*, 71(355):746–751, 1976.
- [44] W.L. Luyben. *Chemical Reactor Design and Control*. John Wiley & Sons, 2007.
- [45] M. Maeder and Y. M. Neuhold. *Practical Data Analysis in Chemistry*. Elsevier, 2006.
- [46] M. Maeder and Y.-M. Neuhold. *Practical Data Analysis in Chemistry*. Elsevier, Amsterdam, NL, 2007.
- [47] E.R. Malinowski. *Factor Analysis in Chemistry*. John Wiley & Sons, New York, USA, 2002.
- [48] W. Marquardt. Model-based experimental analysis from experimental data to mechanistic models of kinetic phenomena in reactive systems. *Chem. Eng. Sci.*, 63:4637–4639, 2008.
- [49] H. Martens and T. Naes. *Multivariate Calibration*. John Wiley & Sons, Chichester, U.K., 2002.
- [50] L. Mátyás, editor. *Generalized Method of Moments Estimation*, volume 5. Cambridge University Press, 1999.
- [51] C. Michalik, M. Brendel, and W. Marquardt. Incremental identification of fluid multi-phase reaction systems. *AIChE J.*, 55:1009–1022, 2009.
- [52] D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, New York, 6th edition, 2005.

- [53] S. Narasimhan and C. Jordache. *Data Reconciliation & Gross Error Detection*. Gulf Professional Publishing, 1999.
- [54] S. Narasimhan and R. S. Mah. Generalized likelihood ratios for gross error identification in dynamic processes. *AIChE J.*, 34(8):1321–1331, 1988.
- [55] Sr. Narasimhan and S. Narasimhan. Data reconciliation using uncertain models. *Int. J. of Advances in Engineering Sciences and Applied Mathematics*, 4(1-2):3–9, 2012.
- [56] B.A. Ogunnaike. *Process Dynamics, Modeling, and Control*. 1994.
- [57] C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained linear state estimation - A moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.
- [58] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for non-linear discrete-time systems: Stability and moving horizon approximations. *IEEE Trans. on Automatic Control*, 48(2):246–258, Feb 2003.
- [59] A. Rasmuson, B. Andersson, L. Olsson, and R. Andersson. *Mathematical Modeling in Chemical Engineering*. Cambridge University Press, 2014.
- [60] C. H. Reinsch. Smoothing by spline functions. *Numerische Mathematik*, 10(3):177–183, 1967.
- [61] R. Rengaswamy, S. Narasimhan, and V. Kuppuraj. Receding-horizon nonlinear Kalman (RNK) filter for state estimation. *IEEE Trans. on Automatic Control*, 58(8):2054–2059, August 2013.
- [62] D. Rodrigues, J. Billeter, and D. Bonvin. Control of reaction systems via rate estimation and feedback linearization. In *12Th International Symposium On Process Systems Engineering (Pse) And 25Th European Symposium On Computer Aided Process Engineering (Escape), Pt A*, volume 37, pages 137–142. Elsevier Science Bv, 2015.
- [63] D. Rodrigues, S. Srinivasan, J. Billeter, and D. Bonvin. Variant and invariant states for reaction systems. *Comp. Chem. Eng.*, 73:23–33, 2015.
- [64] G. Rodriguez-Yam, R. A. Davis, and L. L. Scharf. Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression. *Unpublished manuscript*, 2004.
- [65] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, 44(3):683–700, 2007.

- [66] J. Romagnoli and M. C. Sanchez. *Data Processing and Reconciliation for Chemical Process Operations*, volume 2. Academic Press, 1999.
- [67] D. R. Schneider and G. V. Reklaitis. On material balances for chemically reacting systems. *Chem. Eng. Sci.*, 30:243–247, 1975.
- [68] D.E. Shobrys and D.C. White. Planning, scheduling and control systems: why can they not work together. *Comp. Chem. Eng.*, 24(2):163–173, 2000.
- [69] D. Simon. *Optimal State Estimation: Kalman, H_∞ , and Nonlinear Approaches*. John Wiley & Sons, 2006.
- [70] B. Srinivasan, M. Amrhein, and D. Bonvin. Reaction and flow variants/invariants in chemical reaction systems with inlet and outlet streams. *AIChE J.*, 44(8):1858–1867, 1998.
- [71] S. Srinivasan, J. Billeter, and D. Bonvin. Identification of multiphase reaction systems with instantaneous equilibria. *accepted in Ind. Eng. Chem. Res.*
- [72] S. Srinivasan, J. Billeter, and D. Bonvin. Extent-based incremental identification of reaction systems using concentration and calorimetric measurements. *Chem. Eng. J.*, 207-208:785–793, 2012.
- [73] S. Srinivasan, J. Billeter, S. Narasimhan, and D. Bonvin. Data reconciliation in reaction systems using the concept of extents. *Computer Aided Chemical Engineering (ESCAPE)*, 37:419–424, 2015.
- [74] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2011.
- [75] P. Vachhani, R. Rengaswamy, V. Gangwal, and S. Narasimhan. Recursive estimation in constrained nonlinear dynamical systems. *AIChE J.*, 51(3):946–959, 2005.
- [76] V. Venkatasubramanian, R. Rengaswamy, and S.N. Kavuri. A review of process fault detection and diagnosis: Part ii: Qualitative models and search strategies. *Comp. Chem. Eng.*, 27(3):313–326, 2003.
- [77] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S.N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Comp. Chem. Eng.*, 27(3):293–311, 2003.
- [78] K. V. Waller and P. M. Mäkilä. Chemical reaction invariants and variants and their use in reactor modeling, simulation, and control. *Ind. Eng. Chem. Process Des. Dev.*, 20:1–11, 1981.
- [79] Y. Wang, D.J. Veltkamp, and B.R. Kowalski. Multivariate instrument standardization. *Anal. Chem.*, 63(23):2750–2756, 1991.

-
- [80] J. Workman, B. Lavine, R. Chrisman, and M. Koch. Process analytical chemistry. *Anal. Chem.*, 83(12):4557–4578, 2011.

Curriculum vitae

Sriniketh Srinivasan

Date of birth: 16 June 1990

Email: srinivasan.sriniketh@gmail.com

Education

École Polytechnique Fédérale de Lausanne, Switzerland 2011–2016

PhD candidate in Electrical Engineering

Thesis: *On Decoupling Chemical Reaction Systems*

- *Methods, Analysis and Applications*

Anna University, Chennai, India 2007–2011

B.Tech. in Chemical Engineering

Journal publications

- S. Srinivasan, J. Billeter and D. Bonvin. Sequential model identification of reaction systems - the missing path between the incremental and simultaneous approaches, *Submitted to AIChE J.*
- S. Srinivasan, J. Billeter, S. Narasimhan and D. Bonvin. Shape-constrained data reconciliation of chemical reaction systems, *Submitted to Comp. Chem. Eng.*
- A. Masic, S. Srinivasan, J. Billeter, D. Bonvin and K. Villez. *On the use of shape-constrained splines for biokinetic process modeling*, *Submitted to Comp. Chem. Eng.*
- A. Masic, S. Srinivasan, J. Billeter, D. Bonvin and K. Villez. *Identification of biokinetic models using the concept of extents of reaction*, *Submitted to Env. Sci. Tech.*
- S. Srinivasan, J. Billeter and D. Bonvin. Extent-based identification of multiphase reaction systems with instantaneous equilibria, *Ind. Eng. Chem. Res.*, vol. 55(29), p. 8034-8045, 2016.

- D. Rodrigues, S. Srinivasan, J. Billeter and D. Bonvin. Variant and invariant states for reaction systems, *Comp. Chem. Eng.*, vol. 73, p. 23-33, 2015.
- J. Billeter, S. Srinivasan and D. Bonvin. Extent-based kinetic identification using spectroscopic measurements and multivariate calibration, *Analytica Chimica Acta*, vol. 767, p. 21-34, 2013.
- S. Srinivasan, J. Billeter and D. Bonvin. Extent-based incremental identification of reaction systems using concentration and calorimetric measurements, *Chemical Engineering Journal*, vol. 207-208, p. 785-793, 2012.

Conference Publications

- S. Srinivasan, D. M. Darsha Kumar, J. Billeter, S. Narasimhan and D. Bonvin. On the use of shape constraints for state estimation in reaction systems, 11th IFAC Symposium on Dynamics and Control of Process Systems (DYCOPS), Trondheim (Norway), 2016.
- S. Srinivasan, J. Billeter, S. Narasimhan and D. Bonvin. Data reconciliation in reaction systems using the concept of extents, 12th International Symposium on Process Systems Engineering, Copenhagen (Denmark), 2015.
- N. Bhatt and S. Srinivasan. On cooperative behavior of open homogeneous chemical reaction systems, 1st Indian Control Conference (ICC), Chennai (India), 2015.
- S. Srinivasan, J. Billeter and D. Bonvin. Variant and invariant states for reaction systems. IFAC Workshop on Thermodynamic Foundations of Mathematical Systems Theory (TFMST), Lyon (France), 2013.

