Steinar Kolås

Estimation in nonlinear constrained systems with severe disturbances

PhD-thesis 2008:261

Faculty of Information Technology,
Mathematics, and Electrical Engineering
Department of Engineering Cybernetics

NTNU

Contents

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics, and Electrical Engineering
Department of Engineering Cybernetics
NO-7491 Trondheim
Norway

NTNU Philosophiae Doctor 2008:261
ITK Report 2008:9-W

ISBN 978-82-471-1218-2 (electronic)
ISBN 978-82-471-1218-5 (printed)

Summary

Aluminium is a metal playing an important role in all modern life of today, and it will continue to play an important role in many years to come. As with every business and everyone who participates in the global economy, climate change is a challenge also shared by the aluminium industry. In order for the aluminium industry to meet their goals in reducing greenhouse gas emissions and increase energy efficiency in aluminium production, not only new technology and knowledge will play an important role, but also the speed in which the organization is able to utilize new technology and knowledge is crucial. An important factor for succeeding in reducing greenhouse gas emissions and increase energy efficiency, is e.g. the use of new advanced process control systems and increased process knowledge.

On the way towards new advanced state of the art process control systems, an extension of the current control paradigm may be preferred. In this work, the so-called correlation line, which is very well known in the community of aluminium production, is investigated and discussed, and it is shown that, given constant alumina concentration, the correlation line is unique. Based on the correlation line, a control structure is proposed, aiming for constant mass of aluminium fluoride in the Hall-Heroult cell. It is shown that the control structure reduces process variations and energy consumption without loss of production.

New advanced process control systems imply utilizing state of the art process control systems as e.g. Nonlinear Model Predictive Control (NMPC). There is, however, a growing understanding that the real challenges in advanced model-based control applications like NMPC are the state estimation problem and organizational issues. The state estimation problem - because by not being able to measure all the states, the quality of the estimates determine to a large extent the performance of the NMPC application. Further, the quality of the estimates is dependent of e.g. the accuracy of the model (or system knowledge), the accuracy of the measurements, the estimation algorithm and the tuning of the estimator. Also, but not

that obvious, the knowledge about how the uncertainty/inaccuracy/noise is entering and influencing the system, measurements and control inputs is also of great importance.

One might believe that by defeating these obstacles, the success of an NMPC application is guarantied. However, this might not be the case. Those responsible for commissioning an NMPC application in an organization, e.g. in the process industry, will meet an organization that at the current moment most likely do not have the knowledge to understand what NMPC is all about, which is resistant to changes and might not share the assumptions on which the model are founded. Further, they might not agree in the control actions the NMPC might take, because it does not line up with the knowledge embodied in the organization. These aspects makes the organization the important second challenge, because the organization is the real user of the NMPC application and the degree of success is dependent on their ability to make use of it.

This thesis addresses these two extremities regarding modeling and state estimation by 1) enhancing the toolbox used for modeling and state estimation in constrained nonlinear systems with severe disturbances, and 2) to see the process involved in applying modeling and estimation from an organization point of view.

Item 2) is addressed by looking at possibilities that opens up by the use of modeling and simulation in the context of learning, knowledge creation and knowledge sharing. These issues are addressed by analyzing the role of modeling and simulation in the context of conserving knowledge. The theory is applied to analyze the experience gained from the work processes involved in producing and implementing the content described in the papers Kolås and Støre (2008) and Kolås (2007a). E.g., it is concluded that if the knowledge is complex, the use of modeling and simulation serve as a simplification of complexity, suitable for other parts of the organization that is only interested in the behavior and not the detailed chemistry/physics. In a situation like this, the simplified knowledge could act as a source for learning and knowledge creation.

Item 1) is addressed by comparing the Extended Kalman filter (EKF) and the Unscented Kalman filter (UKF). Several authors have experienced shortcomings applying the EKF to systems with severe nonlinearity and/or constraints. In this work we investigate the use of an alternative to the EKF, the UKF, and a broad overview of different UKF algorithms is given. Further, an extension to the ensemble of UKF algorithms is proposed, and finally, the issue of how to add constraints using the UKF approach is

addressed. The performance of the constrained approach is compared with EKF and a selection of UKF algorithms on nonlinear process systems with multimodal probability density functions. The conclusion is that with an algebraic reformulation of the correction part, the reformulated UKF shows very strong performance on our selection of nonlinear constrained process systems.

Subsequently, noise modeling is studied based on a hypothesis that it is important to model noise correctly. In practice this implies a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty. Alternative concepts of modeling the noise are investigated, and it is shown that modeling noise by introducing it in the system auxiliary variables and control inputs may have a positive impact on estimation performance.

Finally, an algorithm for computing the dynamic eigenvalues for linear time-varying (LTV) systems is proposed. The algorithm can be used to analyze the stability for some types of LTV systems, e.g., for some types of systems, the algorithm could be applied to analyze the stability of the EKF and UKF.

Preface

This thesis is a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Norwegian University of Science and Technology (NTNU).

The research period covers August 2003 to August 2008, and has been accomplished at the Department of Engineering Cybernetics (ITK), NTNU with professor Bjarne A. Foss as supervisor and Senior Scientist Asbjørn Solheim, SINTEF as co-supervisor.

The research was sponsored by Hydro Aluminium AS via the project for developing a Nonlinear Model Predictive Control (NMPC) application for the Hall-Heroult process.

To reach the top of a long winding stair, one has to walk the first step. On the way up, one may open a door to a floor to investigate what is hidden behind. When reaching the top of the stair, looking back, the load carried may not be like one had expected from the start. In analyzing why, one may see that the expectations were not aligned with the real world, and adopting the goals while walking was required.

In the world of math, the essence of the Bayesian approach is to provide a mathematical rule explaining how you should change your existing beliefs in the light of new evidence. In other words, it allows scientists to combine new data with their existing knowledge or expertise (Economist (2000)).

To me this Ph.D. work has become a 'Bayesian stair walk', and the main results presented in this thesis are not as originally planned. One may see it as negative or one may see it as positive. To me it has unquestionably been positive. The main topic was planned to be regarding NMPC. During the first four years of this work, I was, in parallel with doing my Ph.D. research, a project leader for a project with the goal of developing an NMPC application for the Hall-Heroult process. During the progress it became more and more clear that the real challenge in the NMPC project was connected to the estimator, and that the quality of the control was closely related to the quality of the estimates produced by the estimator.

Even though the NMPC project resulted in a patent application, the main focus in this Ph.D. work has been on estimation in constrained systems with severe disturbances. Several doors have been opened and some lead to dead ends, and some not. I think that the most important one opened was the door labeled Unscented Kalman filter (UKF). Combined with appropriate methods for constraint handling, the UKF has shown to be a strong alternative to the Extended Kalman filter.

Acknowledgements

During my work and studies, I was fortunate to interact with a number of people who have offered their advice, assistance, encouragement and inspiration. It is impossible to mention all by names but I feel that the following persons deserve special recognition.

Most of all my thanks go to Johannes Aalbu, Hydro Aluminium. Without his visionary leadership, optimism, encouragement and true believe in the possibilities of new advanced technology mixed with the idea to let 'old horses' gain new knowledge, this work would not have been initiated. His support during the progress of this work is invaluable. A thanks goes also to the leaders in Hydro Aluminium following his position, for supporting the continuation of the work. Also, I am grateful for the support from my colleagues at Hydro Aluminium, and in particular Stefan Lauvli Schley for his valuable feedback during this work.

My thanks goes also to my supervisors, professor Bjarne A. Foss, NTNU and Asbjørn Solheim, SINTEF, for the always good discussions, sharing of ideas and knowledge and for sending me in the right directions when necessary. In that respect also Dr. Tor Steinar Schei, Cybernetica AS deserve special recognition for the good discussions and inputs.

Sincere thanks goes to Dr. Rambabu Kandepu (NTNU), who showed me the direction to the UKF, to Dr. Hardy Shianan (NTNU), Dr. Pieter van der Kloet (Petitrix Research B.V.,The Netherlands) and Dr. F.L.Neerhoff (Neerhoff_productions B.V., The Netherlands) for the good discussions and contributions related to dynamic eigenvalues. I would also like to thank all my Ph.D. colleagues for taking their time to listen to my presentations and in one way or another have helped, inspired and supported me.

I am grateful for the opportunity to have had access to the knowledge network at NTNU regarding my field of interest, and would like to thank all the persons at other institutes that have supported me one way or another.

And last, but not least, my thanks goes to my family and particularly to my wife May-Britt, for her love and kind support. This stair would have

been impossible for me to climb without her by my side.

Publications

Most of the material presented in this thesis has been published, recently been submitted for publication or accepted for publication in journals or conference proceedings. The connection between the publications and the chapters of the thesis is given below.

- Chapter 2 is based on the paper "*Defining and Verifying the 'Correlation Line' in Aluminium Electrolysis*", published in Journal of Metal*, 2007* (Kolås (2007a)).
- Chapter 3 is based on the paper "*Bath Temperature and AlF_3 Control of an Aluminium Electrolysis Cell*", submitted to Control Engineering Practice 2008 (Kolås and Støre (2008)).
- Chapter 4 is an extended version of the abstract "*State estimation IS the real challenge in NMPC*", accepted to conference NMPC'08, Pavia, Italy (Kolås et al. (2008c)).
- Chapter 5 is based on the paper "*Constrained Nonlinear State Estimation based on the UKF approach*", submitted to Computers and Chemical Engineering 2008. (Kolås et al. (2008a)).
- Chapter 6 is based on the paper "*Noise modeling concepts in Nonlinear State Estimation*", submitted to Journal of Process Control 2008 (Kolås et al. (2008b)).
- Chapter 7 is based on the paper "*A recursive algorithm for computing the dynamic eigenvalues*", submitted to Systems & Control Letters 2008 (Kolås et al. (2008d)).
- Chapter 8 (*The use of Modeling and Simulation to Conserve Knowledge*) is unpublished material written in cooperation with associate professor R. Klev, NTNU.

Chapter 1

Introduction

1.1 Aluminium heading for the future

Aluminium is a metal playing an important role in all modern life of today[1]. Aluminium is a unique metal. It is light weight, strong, durable and flexible. It does not rust, is 100% recyclable and can be reused endlessly. It can take many forms and comes in a variety of surface finishes. Its unique combination of properties makes it suitable for many applications. Aluminium helps people and the economy to operate effectively and efficiently, and examples of such includes air, road, rail and sea transport; food and medicine; packaging; construction; electronics and electricity transmission. Aluminium was first produced in large scale in 1888[2], and has become the second most-used metal in the world, after iron. Nearly 75% of all aluminium ever made remains in use today, and combined with a need of only 5% of the energy required for primary production, it represents a growing 'energy and resource bank'. In 2006 the annual primary production of aluminium was around 34 million tons and recycled production around 16 million tons, a total of 50 million tons. In comparisons, 17 million tons of copper, 8 million tons of lead and 0.4 million tons of tin was produced in 2006, which mean that more aluminium is produced than all other non-ferrous metals combined (IAI (2008)).

In a sustainable global economy, the aluminium industry is committed to securing business success and continued growth. This will be achieved

[1] I'll bet a large amount of money on that at least 5 items in your surroundings just now has aluminium as a part of it.

[2] It is common to look at the Hall-Heroult process as it was discovered independently in 1886 by the american Charles M. Hall and the french Paul Heroult. The first large scale aluminium production plant was opened in 1888 in Pittsburgh by Charles M. Hall.

not only by progressively improving its environmental, occupational health and safety performance, but also by increasing its positive socioeconomic contribution through its 'Aluminium for Future Generations' Sustainable Development Programme. This programme for continuous improvement comprises 12 voluntary objectives, covering all key phases of aluminium's life cycle from pre-mining to post-consumer, and is overseen by the International Aluminium Institute (IAI), whose member companies are responsible for over 80% of global aluminium production (IAI (2008)).

As with every business and everyone who participates in the global economy, climate change is a challenge also shared by the aluminium industry. Scientists with the United Nations' Intergovernmental Panel on Climate Change (IPCC) has identified factors that cause unnaturally accelerated rates of global warming, and in this picture the full process of manufacturing new stocks of aluminium is responsible for 1% of the global human-induced greenhouse gas emissions. As the world moves to combat climate change, so do the aluminium industry - by addressing the challenges of climate change by employ a life-cycle approach, focusing on the energy savings to be made through use and reuse as well as on the energy required to produce aluminium products. It is in the use phase that the majority of energy is used and/or saved (e.g. during the useful life of cars, buildings, aircraft, etc.). E.g. related to the car industry, reducing fuel consumption without compromising performance and safety is important, and the high strength-to-weight ratio of aluminium plays a crucial role (IAI (2008)):

- The use of lightweight aluminium components in a vehicle can save six to twelve times the energy taken to produce the primary aluminium used in its construction.
- Up to eight per cent fuel savings can be realized for every 10% reduction in weight.
- One kilogram of aluminium, used to replace heavier materials in a car or light truck, has the potential to eliminate 20kg of CO_2 over the lifetime of the vehicle.
- For other vehicles, such as trains, ferries and aircraft, the potential savings are even greater.

The global aluminium industry has therefore developed a four-pronged voluntary strategy to meet the challenges of climate change, which encompasses the full life-cycle of aluminium from production, to primary use, to recycling and reuse (IAI (2008)):

1. Reduce greenhouse gas emissions from aluminium production.
2. Increase energy efficiency in aluminium production.
3. Maximize used-product collection, recycling and reuse.
4. Promote the light-weighting of vehicles.

In order for the aluminium industry to meet these goals, not only new technology and knowledge will play an important role, but it is believed that also the speed in which the organization is able to utilize the new technology and knowledge is crucial. An important factor for succeeding in reducing greenhouse gas emissions and increase energy efficiency in aluminium production, is the use of new advanced process control systems and increased process knowledge. New advanced process control systems implies utilizing state of the art process control systems as e.g. Nonlinear Model Predictive Control (NMPC). NMPC is not a well defined term, but in our context NMPC means the use of a nonlinear mechanistic model, state estimation, and the solution of an online constrained nonlinear optimization problem.

Challenges in NMPC

There is a growing understanding that the real challenges in advanced model-based control applications like NMPC are the state estimation problem and organizational issues. The state estimation problem - because by not being able to measure all the states, the quality of the estimates determine to a large extent the performance of the NMPC application. Further, the quality of the estimates is dependent of e.g. the accuracy of the model (or system knowledge), the accuracy of the measurements, the estimation algorithm and the tuning of the estimator. Also, but not that obvious, the knowledge about how the uncertainty/inaccuracy/noise is entering and influencing the system, measurements and control inputs is also of great importance.

One might believe that by defeating these obstacles, the success of an NMPC application is guarantied. However, this might not be the case. Those responsible for commissioning an NMPC application in an organization, e.g. in the process industry, will meet an organization that at the current moment most likely do not have the knowledge to understand what NMPC is all about, which is resistant to changes and might not share the assumptions on which the model are founded. Further, they might not agree in the control actions the NMPC might take, because it does not

line up with the knowledge embodied in the organization. These aspects makes the organization the important second challenge, because the organization is the real user of the NMPC application and the degree of success is dependent on their ability to make use of it.

Research objectives for this thesis

This research is guided by the hypothesis that the real challenges in developing and commissioning NMPC applications lies in the state estimation problem and organizational issues. It reflects two extremities regard modeling and (state) estimation. One is to see the process involved in modeling and estimation from an organization point of view, the other is to enhance the toolbox used for modeling and estimation in constrained nonlinear systems with severe disturbances.

An important incentive for using modeling and estimation in the industry is to solve problems where the use of a model and an estimator is believed to produce benefits for the company. However, that requires access to appropriate resources, i.e. people who have sufficient knowledge of the system modeled, as well as mastering the art of modeling and estimation. This may involve substantial investments in both personnel and appropriate equipment. A company entering that route may be vulnerable to employees leaving their position for one reason or another. With this respect one may see the process of modelling and estimation as a process of knowledge creating and sharing where the knowledge produced is conserved in the model, reducing the effect of the knowledge loss related to employees leaving. Based on experience, implementing new technology in a large organization may have challenges of its own. An issue in that respect is that the model may serve as a communication platform easing the transfer of new knowledge into the organization in operation. In Chapter 8 these and other related issues are discussed.

The industry of operating plants based on the Hall-Heroult process have all the characteristics just mentioned. The process is complex and hard to understand, demanding high competence both in operations and in the research area developing new technology. As in all modern industry the organization is always struggling for effectiveness and profitability, and the number of employees may vary with the economic trends also in R&D, making the investment in models an important strategy. Further, new technology related to Hall-Heroult cells means increased current and larger production units (cells). These increases the demand for improved process knowledge (Chapter 2) and improved process control systems (Chapter 3),

and since measurements of important controlled variables are limited, modeling and estimation is a logical route to follow. The Hall-Heroult process is highly nonlinear, and NMPC is by this a natural choice. However, it seems that there is a growing common understanding that the performance of an NMPC application is closely related to the quality of the estimates produced by the applied estimator (Chapter 4, see also Marafioti et al. (2008)). The quality of the estimates may not only be dependent on the accuracy of the model, but also of the estimating method selected (Chapter 5) and how process knowledge is applied (Chapter 6).

1.2 The Hall-Heroult process

The Hall-Heroult process is the dominating process for producing aluminum today (Grjotheim and Kvande (1993)). The fundamentals of the process are to dissolve Al_2O_3 in molten cryolite (also known as electrolyte or bath), and electrically reduce complex aluminum containing ions to pure aluminum. The overall electrochemical reaction is

$$2Al_2O_3 + 3C \rightarrow 4Al + 3CO_2$$

where carbon is fed to the reaction as consumable anodes. By the use of various additives, in particular AlF_3, the operating temperature of the electrolyte can be lowered from 1010^oC to approximately 960^oC. Both decreased temperature and increased excess AlF_3 is believed to be beneficial for the amount of metal produced (current efficiency) and the energy consumption. As molten cryolite is very corrosive, the only component of an acceptable cost presently capable of coexisting with it over time is frozen cryolite. It is therefore necessary to maintain a layer of frozen cryolite (side ledge) to prevent the side walls from eroding. In order to maintain the side ledge there has to be a substantial heat loss through the side ledge and the carbon walls of the cell. The cell voltage applied is typically 4.5V, and the electric current through the cell is typically 150 - 350kA. A sketch of a cell is shown in Figure 1.1. In a modern plant of today 100-300 cells are placed and connected in series. There are three control inputs to the process, anode beam adjustments (controlling energy input), addition of AlF_3 and addition of Al_2O_3, and three controlled variables, electrolyte temperature[3], concentration (or mass) of AlF_3 and concentration of Al_2O_3. A cell is regularly excited since liquid aluminium is tapped, and the anode

[3] Also known as bath temperature.

blocks are changed typically once a day or two each second day. This induces severe disturbances in the energy balance, and it implies that the operating conditions will vary significantly and hence provoke nonlinear cell effects. The process has strong internal couplings, for instance between the mass- and energy balances at the side ledge. The coupled mass and energy balance combined with nonlinear process characteristics and few measurements, makes the Hall-Heroult process challenging to control (Foss and Schei (2005), Drengstig et al. (1998),Gran (1980)).

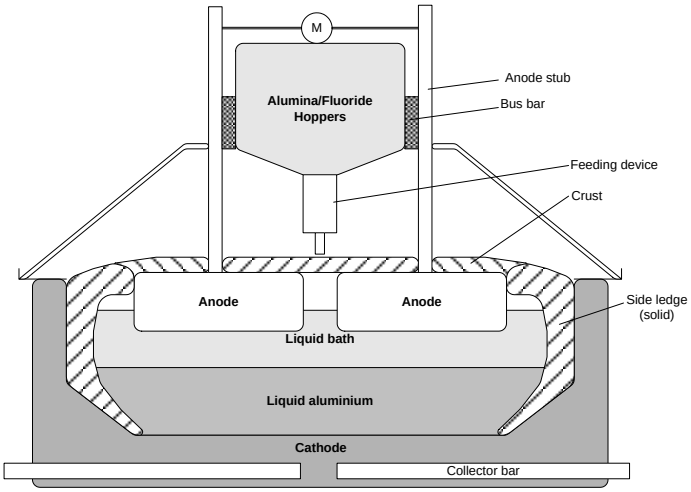

Figure 1.1: The figure illustrates a cell for producing liquid aluminium.

1.2.1 Recent technological review

[4]In the early 1980s, new technology focused on magnetic compensation, point feeders and resistance-tracking process control. The 175 kA barrier was broken, and the new technology enabled operations in a narrower alumina solubility range. The new technology enabled the industry to do bath chemistry modifications that increased the current efficiency and lowered the energy consumption. Typically the energy consumption was between 12.9 DC kWh/kg Al and 13.2 DC kWh/kg Al. The anode-effect frequency was also typically lowered by an order of magnitude, which is a very important environmental factor (greenhouse related).

[4]Mainly from Welch (1999).

By the mid 80's the focus was on 'small' doses point feeders discharging less than 2 kg alumina per dump, and automated systems for distributing the alumina from the large central storage bins to the hoppers on the cell. This focus also had an environmental consequence, namely less dust emissions to the environment.

The trend in the process control system was going from a centralized system to a distributed embedded microprocessor based system, with fast sampling of the line current and the cell voltage, utilizing feeding and energy change according to a predetermined strategy.

Better process control systems and process knowledge in the 1990's made it possible to increase the amperage and the size of the cells. This was preferable because of the economic gains through reduced capital expenditure per unit production.

Moving towards larger cells was not without problem, though. There are a number of issues addressing why the larger cells (greater than 310 kA) do not achieve the low energy consumption of the slightly smaller cells of the early 1980's. These issues includes operation at higher cathode current density; concentration gradients related to less efficient mixing end-to-end in the cell; reduced cell life caused by the more stringent heat balance requirements for freeze protection and the increased cathode corrosion with current density; and the tendency to form sludge because of higher feed frequency and limited mixing. Despite these limitations, the overall economics are more favoured through the reduction of capital cost than through the increase in the energy costs.

1.2.2 The control problem in the Hall-Heroult process

As stated above, the coupled mass and energy balance combined with nonlinear process characteristics and few measurements, makes the Hall-Heroult process challenging to control.

Conventional control

Traditionally the control challenge has been dealt with by designing process control systems that more or less achieve a constant resistance (constant energy input), over/underfeeding of alumina (Al_2O_3) and an aluminium fluoride (AlF_3) feeding strategy based on achieving constant bath acidity. All the control structures are a mix of some types of PID controllers bounded by heuristics (see Stevens Mc Fadden et al. (2001) for further references). In the late 90's, literature addressing the mass AlF_3 instead of

acidity emerged, but still based on some types of PID control (Drengstig (1997), Drengstig et al. (1998)).

Kolås and Støre (2008) propose an alternative method to the common control structure. It utilizes the so called correlation line in order to achieve constant mass of aluminium fluoride in the bath. The use of the correlation line may be seen as a decoupling of the mass and energy balance with respect to aluminium fluoride control. The aluminium fluoride controller is further combined with an energy controller where the energy input to the cell mainly is used to achieve correct bath temperature. A proof of the correlation line as applied in Kolås and Støre (2008) is proposed in Kolås (2007a).

Advanced control

Although the conventional control structures are dominating, several authors have addressed advanced process control structures of control variables regarding the Hall-Heroult process. This includes the adaptive control of alumina addition (Aalbu (1986)), 9 -Box Matrix Control (Rieck et al. (2003)), LQG Control (Stevens Mc Fadden et al. (2006), Gran (1980)), Model Predictive Control (MPC) (Shuiping et al. (2008)) and control structures involving the Neural network approach (Meghlaoui et al. (1998)).

Recently Hydro Aluminium have been active in developing an advanced control structure, by initiating an NMPC project that has resulted in a patent application for NMPC control of the Hall-Heroult process (Kolås (2007b)). The structure of the NMPC controller is depicted in Figure 1.2.

Severe disturbances

Some important characteristics concerned the aluminium electrolysis process is the cyclic change of anodes and cyclic tap of metal. In this industry, most process control system of today does not take into account this cyclic pattern, but strive to achieve a reference resistance, bath temperature and bath acidity. It is believed that there exists a potential in controlling the aluminium electrolysis cell in an optimal manner with regard to the cyclic operations.

There is information available in the system today that is not used for control purposes. This information is the anode position to be changed and the tap schedule from the cast house. If one includes this information in a control strategy, there should be possibilities for further improvements in the performance of an electrolysis cell with respect to productivity and

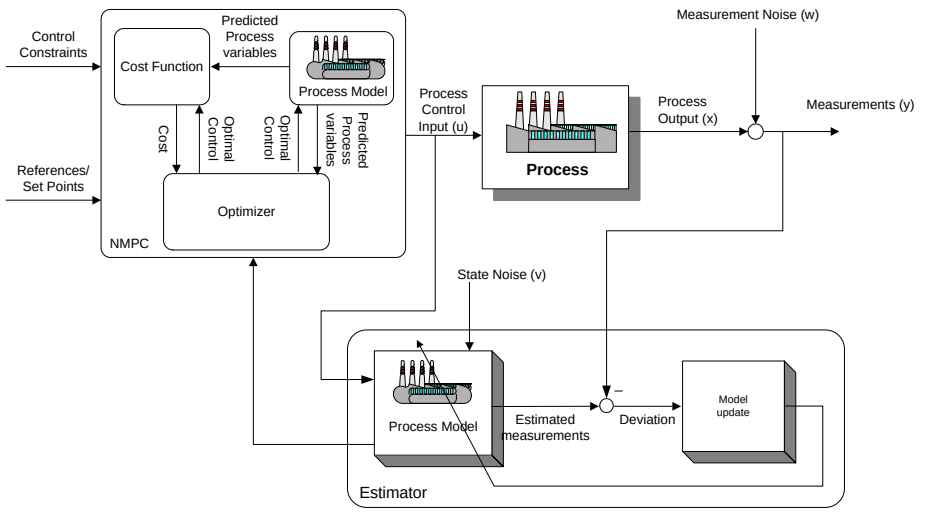

Figure 1.2: The figure indicates the layout of the NMPC controller as applied by Hydro Aluminium AS.

energy savings.

1.3 Estimation in nonlinear constrained systems

An important challenge in an NMPC application is the estimator, in that the complexity and efficiency of the NMPC is closely related to the quality of the estimates produced by the estimator. This is illustrated in Figure 1.3, where data from one of the early tests of NMPC in closed loop control of the Hall-Heroult process is shown.

Figure 1.3 clearly illustrates that the performance of the estimator is crucial for the expected performance of the NMPC application. The quality of the estimates may not only be dependent on the accuracy of the model, but also of the estimating method selected (Chapter 5) and how process knowledge is applied (Chapter 6). This is illustrated in Figure 1.4, where the convergence speed for different estimator algorithms is investigated with respect to erroneous initial values.

Figure 1.4 clearly shows the impact of the estimator algorithms with respect to convergence speed, given erroneous initial values.

These results motivated the further studies of the UKF approach. A broad overview of different UKF algorithms is given in Chapter 5, and

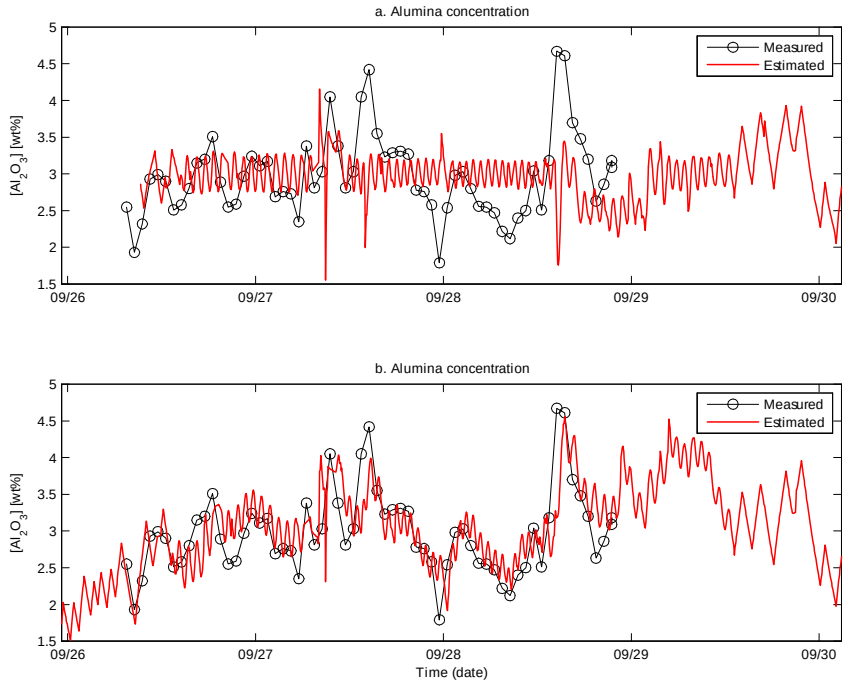

Figure 1.3: The figure shows measured and estimated alumina concentration for different tuning (a) and (b) of an estimator for the Hall-Heroult process. Note that the measured alumina concentration is not available to the estimator.

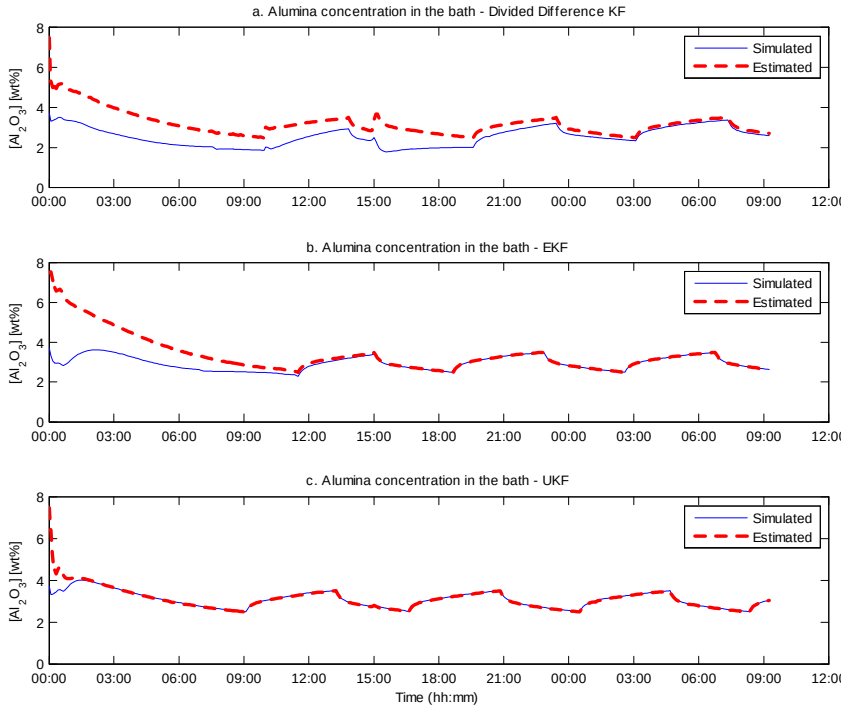

Figure 1.4: The figure illustrates the convergence speed (how fast the estimated curve approaches the simulated curve) for different estimator algorithms with respect to the same erronous initial values. Subplot (a) shows the convergence properties for the applied Divided Difference KF. Subplot (b) show the convergence properties for the applied EKF, and subplot (c) shows the convergence properties for the applied UKF. The investigated case studied is the alumina concentration for the Hall-Heroult process. Note that the estimators do not have identical tuning.

an extension to the ensemble of UKF algorithms is presented. Also the issue of how to add constraints using the UKF approach is addressed. The performance of the constrained approach is compared with EKF and a selection of UKF algorithms on nonlinear process systems with multimodal probability density functions. The conclusion is that with an algebraic reformulation of the correction part, the reformulated UKF shows very strong performance applied to our selection of nonlinear constrained process systems. This work also indicates that adding complexity to the original UKF algorithms have a computational cost (Chapter 5, Appendix B), but it also indicates a potential increase in estimation accuracy (Chapter 5, Appendix B), even though the most complex algorithm not necessarily gives the most accurate estimates. Hence, which algorithm to select is a trial and error process.

1.4 Noise modeling concepts

Recursive state estimation algorithms usually assumes that uncertainty enters through additive white noise sources. Further, unknown and time-varying parameters are often modeled similarly by augmenting the states with a parameter vector. Finally, initial model uncertainty is reflected through the choice of the initial covariance matrices for the states and parameters.

In this work we study noise modeling based on a hypothesis that it is important to model noise correctly. In practice this implies a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty. Alternative concepts of modeling the noise are investigated, and it is shown that modeling noise by introducing it in the system auxiliary variables and control inputs may have a positive impact on estimation performance.

As claimed in the previous chapter, the quality of the estimates may not only depend on the accuracy of the model and of the estimating method selected (Chapter 5), but also of how process knowledge is applied (Chapter 6). One important aspect is how noise/uncertainty is entering the model. A particular challenge with the Hall-Heroult process is that the control inputs may experience severe disturbances, and how this uncertainty is modeled may have effect on the accuracy of the estimates. This is shown in Figure 1.5, where the Hall-Heroult process is simulated in closed loop control by an NMPC application as used by Hydro Aluminium AS. The model used is an adoption of the models found in Saksvikrønning et al. (1976) and Gran (1980). The estimators used in the simulations are the EKF and the UKF.

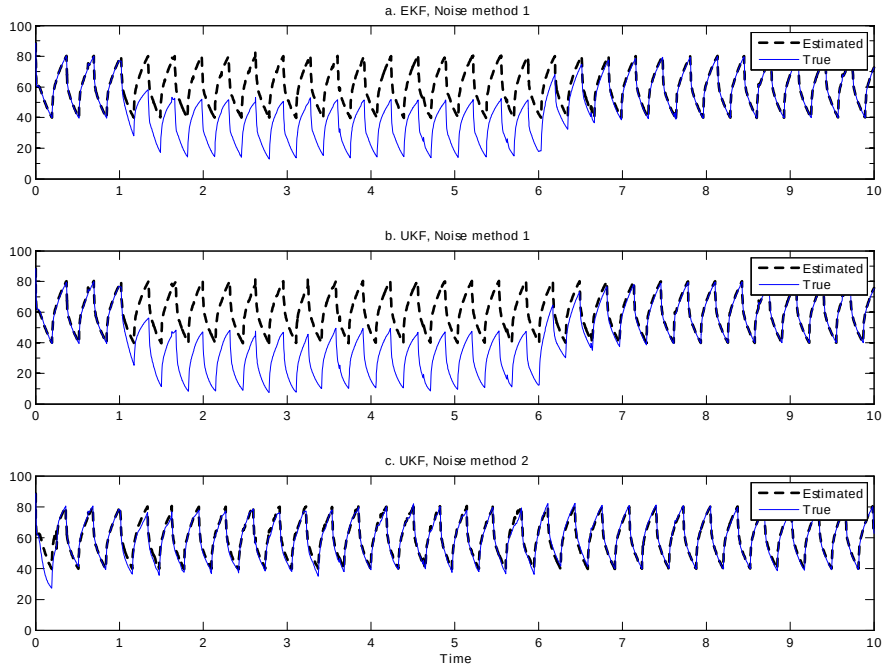

Figure 1.5: The figure indicates the effect of applying different noise conscepts as decribed in Chapter 6 on the Hall-Heroult process.

In the simulation a reduction of 5% in one of the control inputs is introduced in the simulator. The drop in the control input is not known to the control application. The purpose of the simulation is to study the input noise disturbance model described in Chapter 6. Applying the noise modeling concept as described in Chapter 6 regarding input uncertainty, the UKF estimator 'discover' the drop, which is indicated in Figure 1.5 subplot c, by no drop in the **true** state (note that the system is in closed loop control). As indicated by Figure 1.5 subplot a) and b), the drop in the control input is not discovered by the estimators using only the traditional additive noise formulation.

1.5 Dynamic eigenvalues

By applying the Kalman filter (KF) approach to linear time-invariant (LTI) systems, the eigenvalues at time k (λ_k) of the system

$$\lambda_k = eig(A - K_k D) \tag{1.1}$$

tells us if the system converge to the true value or not, and in some cases how fast. Here A is the system matrix, K_k the Kalman gain and D the output (measurement) matrix. In nonlinear systems it is common to use the Extended Kalman filter (EKF). The use of an EKF implies first order linearization resulting in time varying matrices A_k and D_k resulting in a time varying system. However, studying the eigenvalues on the system

$$\lambda_k = eig(A_k - K_k D_k) \tag{1.2}$$

may or may not work. $A_k - K_k D_k$ describe a linear time-varying (LTV) system, and the idea of exploring the dynamic eigenvalues was to investigate if the method, which is valid for linear time varying (LTV) systems, could be used to study stability properties of the EKF. The dynamic eigenvalues cannot be used to determine the stability for LTV systems in a direct fashion. They are, however, closely related to the Lyapunov exponents which in their turn predict the stability for LTV systems. In this work the complex Lyapunov exponent is considered.

Calculating the dynamic eigenvalues for linear time varying (LTV) systems involve solving non-symmetric differential Riccati equations (DRE). Solving these DRE's is challenging, especially when the solution involves singularities. In this work a recursive computer algorithm for computing the dynamic eigenvalues based on the Möbius transform is suggested. Systems with periodic singularities are also considered.

However, requirements on the system formulations structure made the dynamic eigenvalue approach not suitable for further investigation on the convergence performance for the different estimators used on the Hall-Heroult process.

Note that in the work of Xiong et al. (2006) and Xiong et al. (2007) a stability analysis of UKF-based estimators is done by transforming the stability analysis problem into a LTV system. It is believed that the dynamic eigenvalue approach could be used in investigating this LTV system.

1.6 Modeling and conservation of knowledge

The continuous struggle for effectiveness and profitability, facing most industrial organizations, demand frequent structural changes in terms of reorganization, outsourcing or downsizing. This implies that many people will leave their positions in the organization, and important knowledge could be lost. At the same time, recent developments within strategic management and organizational theory argue on the crucial importance of individual and organizational knowledge for growth and development of companies (Hislop (2005), Carlsen et al. (2004), von Krogh et al. (2000), Nonaka and Takeuchi (1995), Spender (1993)). This leaves the company with the combined challenge of preserving and developing its knowledge resources, while at the same time allowing for reducing the number of personnel involved in each work process. In other words, the company has to make crucial knowledge and learning processes less dependent on individuals.

In this work, we will explore the role and possibilities of modeling and simulation in the context of learning, knowledge creating and the sharing of knowledge. Based on a discussion of the nature of knowledge and learning, and the principles of modeling and simulation, as well as the analysis of two practical cases, we suggest that developing mathematical models for computer simulations has a potential for being useful in at least three different forms of knowledge processes:

- It could enhance the organizations capacity to conserve/and or reproduce important knowledge, and thus make a knowledge dependent organization less vulnerable of people leaving their position.
- It could be a tool for knowledge sharing and knowledge creation between people with different areas of expertise, and
- It could be instrumental in creating and communicating simplified and crucial insights into (in principle) processes with high complexity.

In order to substantiate this position, we will present some main positions within the field of knowledge and learning in organizations. Further; relevant topics of modeling and simulation are presented, followed by a discussion where modeling and simulation are put into the context of learning, knowledge creation and knowledge sharing.

1.7 Notation

In this work we address the general continuos nonlinear system given by the state space formulation

$$\begin{aligned} \dot{x}(t) &= f_c(x(t), \theta(t), u(t), v(t)) \\ & \quad x(0) \text{ - given} \end{aligned} \tag{1.3}$$

$$y(t) = h(x(t), w(t)) \tag{1.4}$$

where $x(t)$ denotes the system states, $\theta(t)$ denotes the system parameters, $y(t)$ the output model (or measurements), $u(t)$ control inputs, $v(t)$ system state noise input (stationary or time-varying stochastic), $w(t)$ output (measurement) noise inputs (stationary or time-varying stochastic), $f_c(\cdot)$ and $h(\cdot)$ are nonlinear Lipschitz continous functions. All variables are vectors with appropriate dimensions.

Further we address the general discrete time nonlinear system with constant sampling time (denoted by k) given by the state space formulation

$$\begin{aligned} x_k &= f(x_{k-1}, \theta_{k-1}, u_{k-1}, v_{k-1}) \\ & \quad x_0 \text{ - given} \end{aligned} \tag{1.5}$$

$$y_{k-1} = h(x_{k-1}, w_{k-1}) \tag{1.6}$$

where x_k denotes the system states, θ_k denotes the system parameters, y_k the output model (or measurements), u_k control inputs, v_k system state noise input (stationary or time-varying stochastic), w_k output (measurement) noise inputs (stationary or time-varying stochastic), $f(\cdot)$[5] and $h(\cdot)$ are nonlinear functions.

[5] Note that the nonlinear function $f(\cdot)$ is typically a discretized representation of a nonlinear continous model $f_c(\cdot)$. The discretizition method used is typically Runge-Kutta 4^{th} order.

All variables are vectors with appropriate dimensions. Estimated values are denoted with a $\hat{}$, typically $\hat{x}_k$, $\hat{y}_k$ etc. Further, the simulator and the estimator share the same model, and modeling error may be introduced via θ_k and to some extent through the noise inputs v_k and w_k.

For the case of simplicity regarding the theoretical results we investigate the covariance equation found in the continuos Extended Kalman filter (Henriksen and Balchen (1990))

$$\dot{P} = FP + PF^T + HQH^T - KRK^T \tag{1.7}$$

where all the elements in (1.7) may be time varying. Here F and H denotes the Jacobians found by $F = \frac{\partial}{\partial x} f(x, \theta, u, v)|_{x=\hat{x},\theta,u,v}$, $H = \frac{\partial}{\partial v} f(x, \theta, u, v)|_{x=\hat{x},\theta,u,v}$, Q is the process spectral density and R is the measurement spectral density. Especially the term HQH^T in (1.7) will have our attention in this work, since this is the term that determine how the system noise is injected into the covariance calculation.

We assume without loss of generality, that since the discrete EKF have the same properties as the continuos EKF, the results valid for the continuos EKF also are valid for the discrete EKF, and by that also for all algorithms based on the EKF-framework (as for example the Unscented Kalman Filter (UKF)).

Further, NMPC is not a well defined term in the sense that NMPC may be used for controllers ranging from a slight variation of linear MPC to the online solution of a constrained nonlinear optimization problem. One example of a slight modification to account for nonlinearity is the use of multiple linear models in such a way that the current working point defines which model should be active at a given time instant. Hence, the QP-problem frequently encountered in linear MPC will change as the active model changes. In our context NMPC shall mean the use of a nonlinear mechanistic model, state estimation, and the solution of an online constrained nonlinear optimization problem.

1.8 Contributions of the thesis

The contributions of this thesis are:

1. It is suggested a proof of the so called correlation line regarding the Hall-Heroult process (Chapter 2).

2. It is suggested a control algorithm for bath temperature control and AlF_3 addition regarding the Hall-Heroult process where the correla-

tion line is fundamental (Chapter 3).

3. UKF contributions (Chapter 4 and 5)

   (a) A reformulation of the correction steps.

   (b) Constraints implementation with respect to the reformulated correction steps.

   (c) QP-formulation for calculating the constrained corrected sigma points in the case where the output/measurement model is linear.

   (d) Considerations regarding the choise of square root algorithm in the sigma point calculation.

   (e) Suggested a method of handling singular/negative definit covariance matrices in calculating the sigma points.

4. Noise modeling contributions (Chapter 4 and 6)

   (a) Input uncertainty by modelling it as $u(1 + v_u)$ where u is the input and v_u is the uncertainty.

   (b) Noise in auxiliary variables (an alternative to off-diagonal tuning of the process noise).

   (c) The importance of model accuracy.

5. A recursive algorithm for computing the dynamic eigenvalues (Chapter 7).

6. It is suggested that developing mathematical models for computer simulations has a potential for being useful in at least three different forms of knowledge processes (Chapter 8):

   (a) It could enhance the organizations capacity to conserve/and or reproduce important knowledge, and thus make a knowledge dependent organization less vulnerable of people leaving their position.

   (b) It could be a tool for knowledge sharing and knowledge creation between people with different areas of expertise.

(c) It could be instrumental in creating and communicating simplified and crucial insights into processes with (in principle) high complexity.

1.9 Outline of the thesis

The thesis is divided into three parts, where Part I addresses issues regarding the Hall-Heroult process, Part II addresses constrained nonlinear state estimation and Part III addresses estimation in a knowledge setting.

- **Part I** - Contributions regarding the Hall-Heroult process
  - **Chapter 2** addresses the so-called correlation line theoretically, and by simulations, and verifies the theory by plant data.
  - **Chapter 3** proposes a control structure where the correlation line plays a key role in controlling the bath temperature and the mass of AlF_3 in the Hall-Heroult process.
- **Part II** - Contributions regarding constrained nonlinear state estimation
  - **Chapter 4** is more or less an extract of chapter 4 and 5, but has some elements on Divided Difference Kalman filter not covered by the said chapters.
  - **Chapter 5** addresses the importance of selecting the appropriate estimation algorithm when dealing with estimation in nonlinear systems. The Extended Kalman filter (EKF) and the Unscented Kalman filter (UKF) is compared. Several authors have experienced shortcomings applying the EKF to systems with severe nonlinearity and/or constraints. In this work we investigate the use of an alternative to the EKF, the UKF, and a broad overview of different UKF algorithms is given. Further, an extension to the ensemble of UKF algorithms is proposed, and finally, the issue of how to add constraints using the UKF approach is addressed. The performance of the constrained approach is compared with EKF and a selection of UKF algorithms

on nonlinear process systems with multimodal probability density functions. The conclusion is that with an algebraic reformulation of the correction part, the reformulated UKF shows very strong performance applied to our selection of nonlinear constrained process systems.

- **Chapter 6** studies noise modeling based on a hypothesis that it is important to model noise correctly. In practice this implies a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty. Alternative concepts of modeling the noise are investigated, and it is shown that modeling noise by introducing it in the system auxiliary variables and control inputs may have a positive impact on estimation performance.
- **Chapter 7** propose an algorithm based on the Möbius transform for computing the dynamic eigenvalues for linear time-varying (LTV) systems. The algorithm can be used to analyze the stability for some types of LTV systems. E.g. for some types of systems the algorithm could be applied to analyze the stability of the EKF and UKF.

- **Part III** - Contributions regarding knowledge processes

  - **Chapter 8** addresses modeling and simulation by looking at possibilities that opens up with the use of modeling and simulation in the context of learning, knowledge creation and knowledge sharing. These issues are addressed by analyzing the role of modeling and simulation in the context of conserving knowledge. The theory is applied to analyze the experience gained from the work processes involved in producing and implementing the content presented in Chapter 2 and 3. E.g., it is concluded that if the knowledge is complex, the use of modeling and simulation serve as a simplification of complexity, suitable for other parts of the organization that is only interested in the behavior and not the detailed chemistry/physics. In a situation like this, the simplified knowledge could act as a source for learning and knowledge creation.

– *Enjoy* –

Part I

Contributions regarding the Hall-Heroult process

Chapter 2

Defining and verifying the 'correlation line' in aluminium electrolysis

This chapter is based on (Kolås 2007), as published in Journal of Metal, 2007. It is an advantage to be familiar with the Hall Heroult process in order to fully understand the content of this chapter.

2.1 Abstract

Within the community of the aluminium electrolysis industry it is well known and well documented that there exists a relationship between bath temperature and excess AlF_3. This relationship is often referred to as the correlation line. Inspired by this awareness, this paper proves the correlation line theoretically. This is done in two steps. First, the correlation line is derived from a model of mass and energy balance of an electrolysis cell. Second, the theoretically derived correlation line is verified experimentally.

Keywords: Metal industry, Aluminium electrolysis, Correlation line, Estimation.

Nomenclature

$\cdot$	Rate of change	
[] or C_{xx}	Concentration	$[wt\%]$
$\hat{}$	Estimated values	
ACD	Anode-Cathode Distance	$[m]$
A_{sw}	Side wall area	$[m^2]$
H_a	Heat transfer coefficient, ambient	$[W/(m^2{}^\circ C)]$
h_b	Bath height	$[m]$
h_m	Metal height	$[m]$
H_{sl}	Heat transfer coefficient, side ledge	$[W/(m^2{}^\circ C)]$
h_{sl}	Specific enthalpy, side ledge	$[J/kg]$
I	Line current	$[kA]$
k_0	Constant in the correlation equation	$[^\circ C]$
k_1	Constant in the correlation equation	$[^\circ C/wt\%AlF_3]$
k_2	Constant in the correlation equation	$[^\circ C/wt\%Al_2O_3]$
k_{sl}	Thermal conductivity of the side ledge	$[W/(m^\circ C)]$
m_{xx}	Mass	$[kg]$
Q_{sl}	Heat flow into the side ledge	$[W]$
Q_{sw}	Heat flow through the side wall	$[W]$
R_b	Pseudo resistance	$[\mu\Omega]$
SH	Superheat	$[^\circ C]$
T_0	Ambient temperature	$[^\circ C]$
T_b	Bath temperature	$[^\circ C]$
T_{liq}	Liquidus temperature	$[^\circ C]$
U_{Cell}	Cell voltage	$[V]$
U_{ext}	A constant between 1.5-1.8	$[V]$
x_{sl}	Side ledge thickness	$[m]$
ρ_{sl}	Density of the side ledge	$[kg/m^3]$

2.2 Introduction

It is well known in the aluminium community that both AlF_3 additions and the bath temperature have an influence on the acidity[1] due to variation in side ledge thickness. The relationship between the bath temperature and the acidity is referred to as bath temperature–acidity correlation, or simply the correlation line. In Desclaux (1987) and Salt (1990) this is thoroughly discussed based on achieved measurements showing the correlation of bath temperature and acidity.

[1] Acidity is also referred to as excess AlF_3.

The scope of this paper is to prove the correlation line theoretically, and to enlighten the reader unfamiliar with the aluminium electrolysis process, a system description is presented in chapter 3.3. Further, the relevant theoretical background is presented in chapter 3.4 and the results in chapter 3.5. Finally, the results are discussed in chapter 3.6.

2.3 System description

2.3.1 The Hall-Heroult process

The Hall-Heroult process[2], named after its inventors, is the only method by which aluminium is produced industrially today. Liquid aluminium is produced by the electrolytic reduction of alumina (Al_2O_3) dissolved in an electrolyte, referred to as bath, which mainly consists of cryolite (Na_3AlF_6). A sketch of the alumina reduction cell is shown in Figure 2.1.

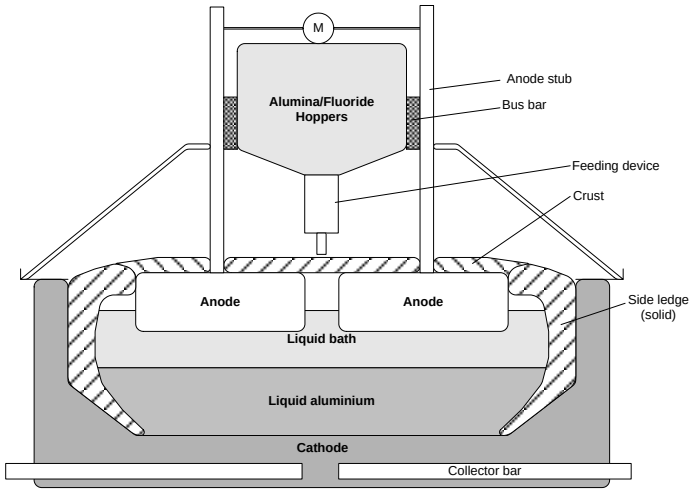

Figure 2.1: A sketch of the main features of an alumina reduction cell (Prebake).

In the alumina reduction cell, hereafter referred to as the cell, one (Søderberg) or several (Prebake) carbon anodes are dipped into the bath. The alumina is consumed electrochemically at the anode (Solheim 2005),

[2]The material presented regarding the Hall-Heroult process is mainly taken from Grjotheim and Kvande (1993).

$$\frac{1}{2}Al_2O_3 + 3NaF + \frac{3}{4}C \rightarrow AlF_3 + \frac{3}{4}CO_2 + 3Na^+ + 3e \qquad (2.1)$$

As can be seen from (2.1), the carbon anode is consumed during the process (theoretically 333 kg C/t Al).

The lower part of the cell, the cathode, consists of a steel shell lined with refractory and thermal insulation. A pool of liquid aluminium is formed on top of the carbon bottom. The cathode, in the electrochemically sense, is the interface between the liquid aluminium and the bath,

$$AlF_3 + 3Na^+ + 3e \rightarrow Al + 3NaF \qquad (2.2)$$

and the total cell reaction becomes

$$\frac{1}{2}Al_2O_3 + \frac{3}{4}C \rightarrow Al + \frac{3}{4}CO_2 \qquad (2.3)$$

Pure cryolite (Na_3AlF_6) has a melting point of 1011 °C. To lower the melting point, the liquidus temperature, aluminium fluoride (AlF_3) and calcium fluoride (CaF_2), to mention the most important ones, are added to the bath. The bath composition in a cell may typically be 6-13 [wt%] AlF_3, 4-6 [wt%] CaF_2, and 2-4 [wt%] Al_2O_3. Lowering the liquidus temperature makes it possible to operate the cell at a lower bath temperature, but at the expense of reduced solubility of Al_2O_3 in the bath, demanding good Al_2O_3 control. It should be mentioned that if the concentration of Al_2O_3 gets too low (less than approx. 1.8 wt%), the cell enters a state called anode effect. During anode effect, the cell voltage increases from the normal 4-4.5 V up to 20-50 V. Anode effect is a highly unwanted state, not only because it represents a waste of energy and a disturbance of the energy balance, but also because strong greenhouse gases (CF_4 and C_2F_6) are produced at the anode (Solheim 2005). Very often the anode effect requires an intervention of an operator.

The bath temperature during normal cell operation is between 940 °C and 970 °C.

The bath is not consumed during the electrolytic process, but some is lost, mainly during vaporization. The vapor mainly consists of $NaAlF_4$. In addition, some bath is lost by entrainment of small droplets, and water present in the alumina feed and air sucked into the cells reacts to form HF (Solheim (2005)). In order to protect the environment, the gas is collected and cleaned in a scrubbing system. More than 98% (Hyland et al. (2001)) of the AlF_3 is recovered in the scrubbing system and recycled back to

the cells. In addition, the content of Na_2O and CaO in the fed Al_2O_3 neutralize AlF_3 (Hyland et al. (2001)). The neutralization is also a function of the penetration of sodium into the cathode, and hence, the cell age. As an example, a 170 kA cell emits about 60 equivalent kg AlF_3 per. 24h, and uses approximately 2500 kg Al_2O_3 per. 24h. The amount of AlF_3 due to neutralization for a 170 kA cell is between 0 and 20 kg per. 24h (dependent of cell age). However, since most of the AlF_3 is recycled, the real consumption of AlF_3 is very small compared to the consumption of Al_2O_3.

At the sidewalls of the cathode there is a frozen layer, called side ledge, which protects the sidewall from erosion.

The composition of the side ledge is mainly pure Na_3AlF_6 with some CaF_2 (Thonstad and Rolseth (1983)). The thickness of the side ledge is a function of the heat flow through the sides, which is a function of the difference in bath temperature and liquidus temperature . Since it is assumed that the side ledge composition is mainly Na_3AlF_6, this means that the total mass of cryolite in the bath varies, while the masses of AlF_3 and Al_2O_3 do not vary with the side ledge thickness. Further, since the concentration of an additive is the mass of the additive divided by the total mass of bath, the variation in the side ledge thickness introduces variation in the concentrations. Hence, the changes in the concentrations introduce changes in the liquidus temperature, which introduces changes in the superheat, affecting the side ledge thickness.

The challenge is thereby to ensure stable cell operations resulting in a stable protective side ledge, while minimizing energy input and maximizing production.

2.3.2 Process control

In controlling a cell, there are three main controlled variables: bath temperature, concentration of AlF_3 and concentration of Al_2O_3, and three control inputs: anode beam adjustments (controlling energy input), addition of AlF_3 and addition of Al_2O_3.

The dynamics in reducing the mass of AlF_3 is slow, and the control of the concentration of AlF_3 has to deal with slow responses when changing the AlF_3 concentration (Meghlaoui et al. (2002), Entner and Gudmundsson (1996), Entner (1995), Entner (1993) , Entner (1992), Drengstig et al. (1998), Wilson (1992), Haupin and Kvande (1993) , Taylor (1992) , Peyneau (1988), Drengstig (1997)). The dynamics in the mass of Al_2O_3 is fast, and the control of the concentration of Al_2O_3 has to deal with quick responses.

The control of the concentration of Al_2O_3 is usually considered as an isolated problem.

The bath temperature is usually measured manually once a day or at least once a week. The concentration of AlF_3 (acidity) is typically measured manually once or twice a week, while the concentration of Al_2O_3 is not normally measured at all, only in conjunction with experiments.

The only continuous measurements is the bath pseudo resistance R_b defined as

$$R_b = \frac{U_{cell} - 1.7}{I} \qquad [\mu\Omega] \tag{2.4}$$

R_b is used as an input for the anode beam adjustment, and acts as a control variable in conjunction with the energy input to the cell.

Because the energy balance and the mass balance are coupled through the side ledge (see for example (Drengstig, 1997, Ch. 5)), the control of a cell must be considered as a non-linear multivariable control problem.

2.3.3 Motivation

The relationship between the bath temperature and the acidity is well known and documented within the primary aluminium community (e.g. Desclaux (1987), Salt (1990), Chen and Taylor (2005), Kolås and Støre (2008)). Even though this relationship is well known, to my knowledge and research, there exists no theoretical proof of it. The motivation of this work is to prove the relationship.

2.3.4 Experimental conditions

All the results referred to in this paper are from experiments done on end-to-end cells (177 kA) at Hydro Aluminium AS.

2.4 Theoretical background

2.4.1 Mass balance & liquidus temperature

A phase diagram of the system ($NaF - AlF_3$), in which the liquidus temperature is a key figure, is sketched in Figure 2.2

There exist several versions of the equation for the liquidus temperature, and one is described by (Solheim and Støen (2005)):

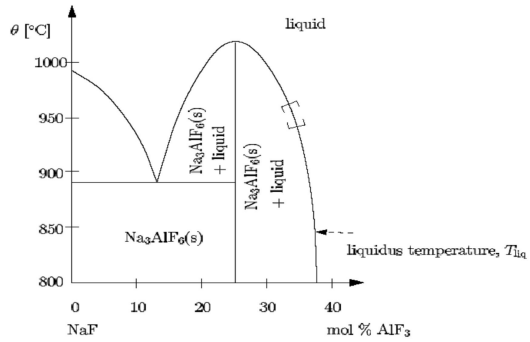

Figure 2.2: A sketch of the phase diagram in the system NaF – AlF_3. The typical operating regime is indicated with brackets (Drengstig 1997).

$$
\begin{aligned}
T_{liq} = \; & 1011 + 0.63\ [AlF_3] - 0.135\ [AlF_3]^{2.2} \\
& - \frac{\left(8.1+2.2e^{-5}[AlF_3]^{3.5}[Al_2O_3]\right)[Al_2O_3]}{1+0.119[Al_2O_3]-0.012[Al_2O_3]^{1.5}} \\
& - \frac{2.55[CaF_2]}{1+0.003[CaF_2]} \\
& + 0.0178\,[AlF_3]^2\left(1 - e^{-(2-0.08[AlF_3])[CaF_2]}\right) \\
& - 0.0012\,([CaF_2]\,[AlF_3])^2 \\
& - \frac{9.05[AlF_3]}{1+0.00562[LiF]+0.0016[CaF_2]^2} \\
& - 0.3\,([LiF]\,[CaF_2])^{0.5}\left(1 - 1e^{-5}\,[AlF_3]^4\right) \\
& - 3.95\,[MgF_2] \\
& - 3.95\,[KF]
\end{aligned}
\qquad [^oC] \qquad (2.5)
$$

The liquidus equation is multidimensional and describes a non-linear multidimensional plane. By varying the concentration of Al_2O_3 and AlF_3, while keeping the other concentrations constant, (2.5) represents a three dimensional plane (Figure 2.3).

By linearizing (2.5) around the point $(T_0, [AlF_3], [Al_2O_3]) = (962, 10, 2.5)$, varying the concentration of Al_2O_3 between 2.5 and 5 $wt\%$, and the concentration of AlF_3 between 10 and 13 $wt\%$ while keeping the other concentrations constant, we get:

$$\hat{T}_{liq} = k_0 + k1\,([AlF_3] - 10) + k_2\,([Al_2O_3] - 2.5) \quad [^oC] \qquad (2.6)$$

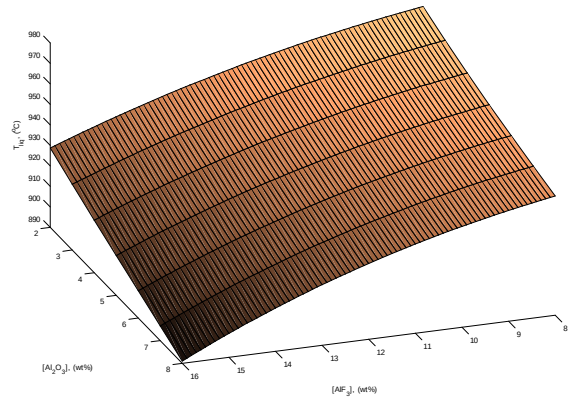

Figure 2.3: The liquidus plane when varying $[Al_2O_3]$ and $[AlF_3]$

where[3] $k_0 = 964.4$ $[°C]$, $k_1 = -5.6$ $[°C/wt\%AlF_3\ ]$ and $k_2 = -5.2$ $[°C/wt\%Al_2O_3\ ]$.

The deviation between (2.5) and (2.6) is illustrated in Figure 2.4.

2.4.2 Energy balance & liquidus temperature

In steady state, the heat flux into the side ledge is in balance with the heat flux through the sidewall (Thonstad and Rolseth (1983)). The heat flux through the side is illustrated in Figure 2.5.

The heat flux into the side ledge is described by

$$Q_{sl} = H_{sl} A_{sw} \left(T_b - T_{liq}\right) \qquad [W] \tag{2.7}$$

The heat flux through the sidewall is described by

$$Q_{sw} = \frac{A_{sw}}{\frac{1}{k_{sl}} x_{sl} + \frac{1}{H_{sw}} + \frac{1}{H_a}} \left(T_{liq} - T_0\right) \qquad [W] \tag{2.8}$$

When not in steady state, the dynamic behavior of the side ledge x_{sl} can be described by:

$$\dot{x}_{sl} = \frac{dx_{sl}}{dt} = -\frac{Q_{sl} - Q_{sw}}{h_{sl} \rho_{sl} A_{sw}} \qquad [m/s] \tag{2.9}$$

[3] k_0 in (Kolås 2007) is 962.

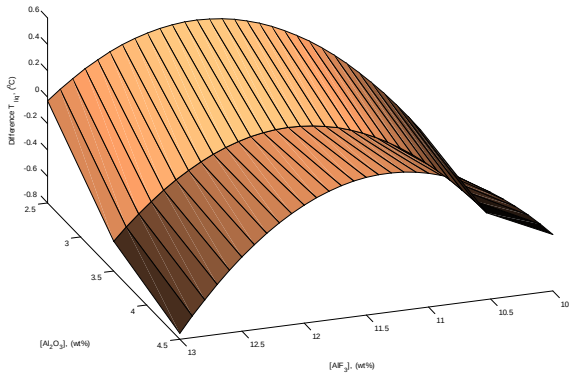

Figure 2.4: The deviation between the full liquidus equation (2.5) and the simplified liquidus equation (2.6).

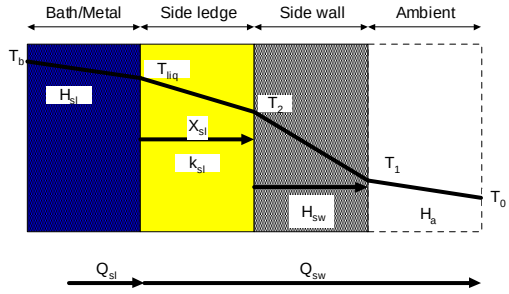

Figure 2.5: The heat balance through the side.

By defining

$$f\left(x_{sl}\right)=\frac{1}{k_{sl}}x_{sl}+\frac{1}{H_{sw}}+\frac{1}{H_a} \tag{2.10}$$

and using (2.7) and (2.8) to solve (2.9) with respect to T_{liq} we get

$$T_{liq}=\frac{H_{sl}f\left(x_{sl}\right)}{1+H_{sl}f\left(x_{sl}\right)}T_b+\frac{T_0}{1+H_{sl}f\left(x_{sl}\right)}+\frac{h_{sl}\rho_{sl}f\left(x_{sl}\right)}{1+H_{sl}f\left(x_{sl}\right)}\dot{x}_{sl} \quad [^oC] \tag{2.11}$$

In steady state, i.e. $\dot{x}_{sl}=0$, (2.11) becomes

$$T_{liq}=\frac{H_{sl}f\left(x_{sl}\right)}{1+H_{sl}f\left(x_{sl}\right)}T_b+\frac{T_0}{1+H_{sl}f\left(x_{sl}\right)} \quad [^oC] \tag{2.12}$$

2.4.3 Coupled mass & energy balance

As mentioned in the introduction, the mass and energy balance are coupled through the side ledge. One of the key elements in the coupling is the liquidus temperature of the bath. Since the liquidus temperature must be on the liquidus plane, (2.5) and (2.12) must describe the same phenomena. By using the linearized liquidus equation (2.6) in (2.11) we get

$$k_0+k1\left(\left[AlF_3\right]-10\right)+k_2\left(\left[Al_2O_3\right]-2.5\right)=\begin{array}{c}\frac{H_{sl}f(x_{sl})}{1+H_{sl}f(x_{sl})}T_b\\+\frac{T_0}{1+H_{sl}f(x_{sl})}\\+\frac{h_{sl}\rho_{sl}f(x_{sl})}{1+H_{sl}f(x_{sl})}\dot{x}_{sl}\end{array} \tag{2.13}$$

Solving (2.13) for $[AlF_3]$ we get

$$\left[AlF_3\right]=f_1\left(x_{sl}\right)T_b+f_2\left(x_{sl}\right)+f_3\left(x_{sl}\right)\left[Al_2O_3\right]+f_4\left(x_{sl}\right)\dot{x}_{sl} \tag{2.14}$$

where

$$f_1\left(x_{sl}\right)=\frac{1}{k_1}\frac{H_{sl}f\left(x_{sl}\right)}{1+H_{sl}f\left(x_{sl}\right)} \quad [wt\%/^oC] \tag{2.15}$$

$$f_2\left(x_{sl}\right)=\frac{1}{k_1}\frac{T_0}{1+H_{sl}f\left(x_{sl}\right)} \quad [wt\%] \tag{2.16}$$

$$f_3(x_{sl}) = -\frac{k_2}{k_1}([Al_2O_3] - 2.5) - \frac{k_0}{k_1} + 10 \quad [wt\%] \tag{2.17}$$

$$f_4(x_{sl}) = \frac{1}{k_1}\frac{h_{sl}\rho_{sl}f(x_{sl})}{1 + H_{sl}f(x_{sl})} \quad [wt\%\ s/m] \tag{2.18}$$

with the steady state solution

$$[AlF_3] = f_1(x_{sl})T_b + f_2(x_{sl}) + f_3(x_{sl})[Al_2O_3] \quad [wt\%] \tag{2.19}$$

It is evident that f_1, f_2 and f_4 are non-linear functions, while f_3 is a constant only dependent on $[Al_2O_3]$ when assuming k_1 and k_2 are constants.

2.5 Results

2.5.1 Simulation

A model based on the equations above was simulated using Matlab[4]. The assumptions made were:

1. The bath and metal height was perfectly controlled to a constant level,
2. Constant $[Al_2O_3]$, and
3. Perfect bath temperature control.

Proper action in the model was taken to achieve this. Further, the bath temperature was taken to vary from 960°C down to 945.5°C in steps of –2.5°C/24h. The reason why the model was perturbed within this small temperature range only, is that the assumption of constant $[Al_2O_3]$ constrains the side ledge thickness, and hence the temperature range where the model is physically valid. At 945.5°C it was added 252 kg AlF_3, and the bath temperature was raised to 958°C in steps of 2.5°C/24h.

Before the simulation results are discussed, it is appropriate to look at the functions f_1, f_2, f_3 and f_4 (Equations (2.15)-(2.18)) as functions of time (Figure 2.6).

[4]Note that the full model is not covered in this chapter. The state vector in the simulations contains the mass of Al_2O_3 ($m_{Al_2O_3}$), mass of AlF_3 (m_{AlF_3}), mass of bath (m_b) and the side ledge thickness (x_{sl}).

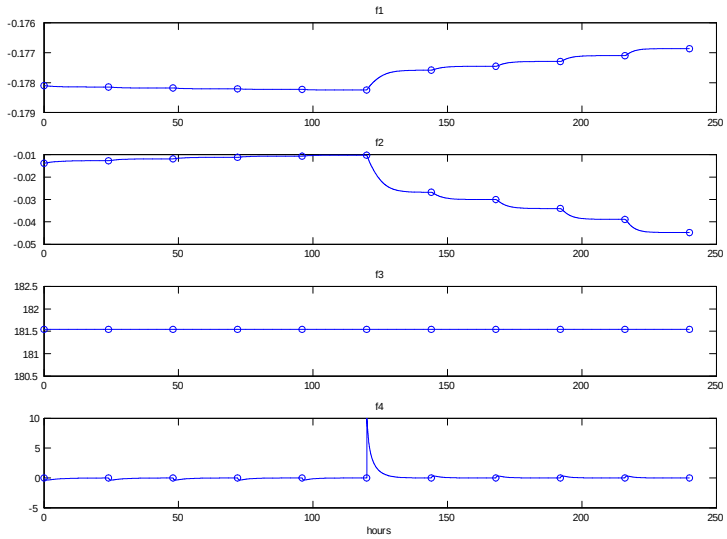

Figure 2.6: The functions f_1, f_2, f_3 and f_4 vs. time

As Figure 2.6 shows, the term f_1 ((2.15)) varies due to variations in the side ledge thickness. The variations is between –0.1782 and -0.1781 before the addition of AlF_3, and between –0.1776 and -0.1769 after the addition. The term f_2 (2.16) varies between –0.0138 and –0.0103 before the addition of AlF_3, and between –0.0447 and –0.0268 after the addition. The term f_3 ((2.17)) is constant during the simulation, due to constant $[Al_2O_3]$ and constant parameterization of (2.5). The term f_4 ((2.18)) does not influence the steady state solution of (2.19), and will be no further mentioned in this text.

The results of the simulation with respect to start and end conditions for the cell are shown in Table 7.2.

The steady state values of the bath temperature and $[AlF_3]$ as functions of time are shown in Figure 2.7 below.

It is also worth noticing that the deviation between the estimated $[AlF_3]$ (2.19) and the simulated $[AlF_3]$ (mass balance) is very small.

The correlation lines between bath temperatures and $[AlF_3]$ before and after the addition of AlF_3 are shown in Figure 2.8.

The first order correlation line before the addition of AlF_3 becomes:

$$\hat{C}_{AlF_3} = -0.169T_b + 172.83 \quad [wt\%] \tag{2.20}$$

Table 2.1: Initial conditions and final conditions of the simulation.

Para	*Start*	*End*	*Change*	*Unit*
[AlF 3]	10,50	11,73	1,23	*wt%*
[Al 2 O 3]	3,32	3,32	0,00	*wt%*
Tb	960,00	960,00	0,0000	*°C*
hb	20,00	20,00	0,00	*cm*
hm	20,00	20,00	0,00	*cm*
Tliq	957,11	950,63	-6,48	*°C*
SH	2,89	9,37	6,48	*°C*
Xsl	19,760	3,782	-15,98	*cm*
mb	5259,45	6855,07	1595,62	*kg*
mAlF 3	552,24	804,24	252,00	*kg*
mAl 2 O 3	174,61	227,59	52,97	*kg*
Asw	11,43	12,06	0,64	*m^2*
Qsl	23,12	79,10	55,98	*kW*
Qsw	23,12	79,10	55,98	*kW*
ACD	41,48	40,62	-0,86	*mm*

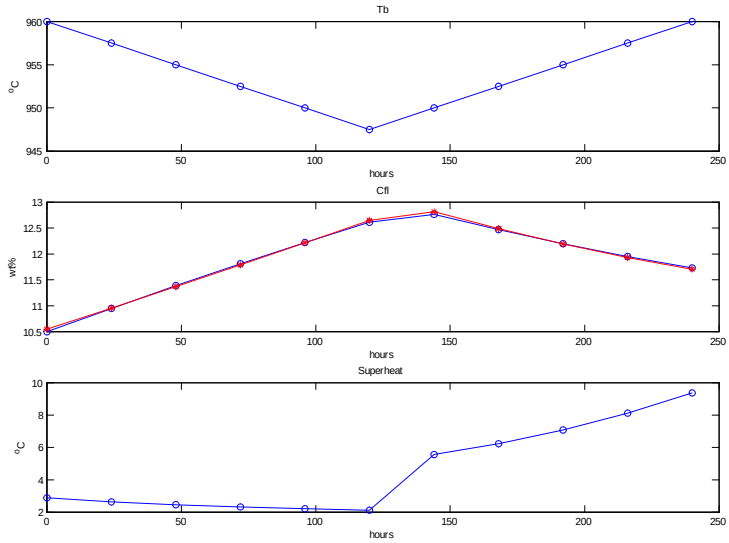

Figure 2.7: The steady state results from the simulation study. The uppermost graph is the simulated bath temperature (T_b). The middle graph shows the simulated acidity ($m_{AlF_3}/m_b \cdot 100$) and estimated acidity ($\hat{m}_{AlF_3}/\hat{m}_b \cdot 100$) (almost identical). The lower graph shows the simulated super heat ($T_b - T_{liq}$).

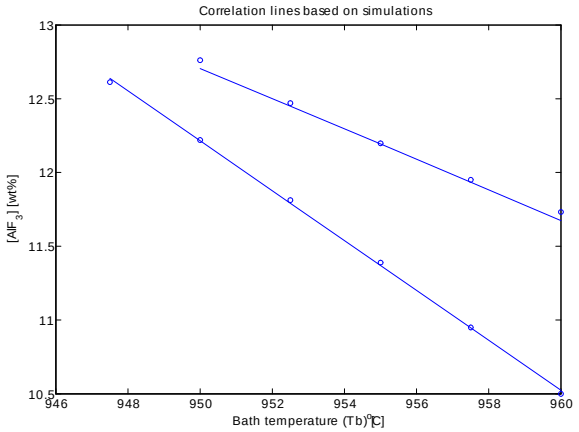

Figure 2.8: The correlation lines before (lower line) and after (upper line) the addition of AlF_3

The first order correlation line after the addition of AlF_3 becomes

$$\hat{C}_{AlF_3} = -0.103T_b + 110.71 \quad [wt\%] \tag{2.21}$$

2.5.2 Experimental

252 kg AlF_3 was added to a 177 kA prebake cell, resulting in the correlation lines shown in Figure 2.9.

The first order correlation line before the addition of AlF_3 becomes:

$$\hat{C}_{AlF_3} = -0.187T_b + 190 \quad [wt\%] \tag{2.22}$$

The first order correlation line after the addition of AlF_3 becomes

$$\hat{C}_{AlF_3} = -0.119T_b + 127 \quad [wt\%] \tag{2.23}$$

2.6 Discussion

(2.19) shows that the correlation line is a function of the side ledge thickness and the $[Al_2O_3]$. Also (2.19) indicates that if we (hypothetically) were able to keep the side ledge thickness constant, the offset is a function of the

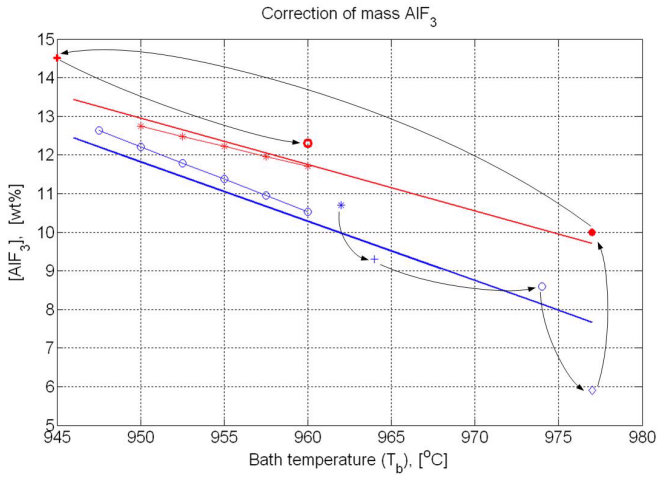

Figure 2.9: The response of addition of AlF_3 with respect to changes in correlation lines. The lower solid line is the correlation line based on measurements prior to the addition of AlF_3, and the upper solid line is the preferred correlation line. The lines with circles are the lines from the simulation above. The arrows indicate the time response of the AlF_3 addition in operations. The addition of AlF_3 was executed at the tip of the third arrow. After the addition, the next three measurements indicate a lifted correlation line.

$[Al_2O_3]$ only. By increasing the $[Al_2O_3]$, the correlation line is lowered and by decreasing the $[Al_2O_3]$, the correlation line is raised.

Further, (2.19) shows that by keeping the $[Al_2O_3]$ constant, the correlation line is determined by the side ledge thickness. The side ledge thickness at a given bath temperature is determined by the liquidus temperature. Since the $[Al_2O_3]$ is constant, the mass of AlF_3 is determining the liquidus temperature (assuming all other concentrations constant). Hence, a constant mass of AlF_3 and a constant $[Al_2O_3]$ determines the correlation line (assumed that all other concentrations are constant).

Simulations indicate that if the $[Al_2O_3]$ is kept constant, an addition of AlF_3 is able to achieve a sufficient different correlation line. This is shown in Figure 2.8, and is to some extent verified by experiments (Figure 2.9).

In the experiment, it was assumed that the alumina controller kept a constant $[Al_2O_3]$. Figure 9 shows that the last point (T_b, $[AlF_3]$) at $T_b = 960$ $^\circ C$ lays both above the simulated result and the preferred correlation line. By assuming that the $[Al_2O_3]$ dropped 1 wt% when adding 252 kg AlF_3, the simulated correlation line at $T_b = 960$ °C is closer to the correlation line achieved by the experiments. The simulated correlation line adjusted by a change in $[Al_2O_3]$ is shown in Figure 2.10.

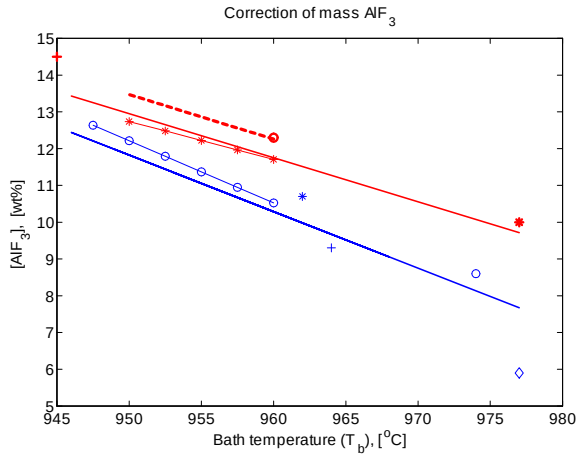

Figure 2.10: This is the same figure as figure 2.9, but where the bold dotted line is the simulated correlation line when also assumed a drop in the $[Al_2O_3]$.

Since the $[Al_2O_3]$ was not measured during the experiment, the assump-

tion that there was a drop in the $[Al_2O_3]$ is pure speculation, but simulation shows that it is likely.

2.7 Conclusions

In this paper we have proved the correlation line theoretically. By (2.19) it is evident that the offset in the correlation line can be adjusted by the $[Al_2O_3]$. Further it is evident that if the $[Al_2O_3]$ is constant, the correlation line is unique when assuming steady state and constant mass of AlF_3. But that does also mean that if the mass of AlF_3 varies and/or the $[Al_2O_3]$ varies, the correlation line varies. In that case we have infinitely many correlation lines.

2.8 Acknowledgement

The author acknowledge the following persons for the support during the writing of this paper: Dr. T.Støre and Dr. H.Kvande at Hydro Aluminium AS, A.Solheim at SINTEF, Dr. M.Hillestad at Cybernetica A/S and Professor B.A.Foss at NTNU.

Chapter 3

Bath temperature and AlF_3 control of an aluminium electrolysis cell

This chapter is based on (Kolås and Støre 2008) as submitted[1] *to Control Engineering Practice. It is an advantage to be familiar with the Hall Heroult process in order to fully understand the content of this chapter.*

3.1 Abstract

While the control of alumina in an aluminium electrolysis cell is considered mature, the control of the energy input and the mass of AlF_3 is done in almost as many ways as there are aluminium smelters. By designing a control system based on plant and process knowledge, it is possible to achieve a reduction in the variation of the bath temperature and the mass of aluminium fluoride in an aluminium electrolysis cell. By reducing the variations it is easier to achieve the desired target and therefore more stable operation and reduced energy consumption without loss of production.

Keywords: Metal industry, Aluminium electrolysis, Process control, Bath temperature control, AlF_3 control, Estimation, Decoupling.

[1] Resubmitted version based on feedback from reviewers.

Nomenclature

$\dot{\cdot}$	Rate of change	
$\hat{}$	Estimated values	
ACD	Anode-Cathode Distance	$[m]$
A_{sw}	Side wall area	$[m^2]$
C_{xx}	Concentration	$[wt\%]$
H_a	Heat transfer coefficient, ambient	$[W/(m^{2\circ}C)]$
h_b	Bath height	$[m]$
h_m	Metal height	$[m]$
H_{sl}	Heat transfer coefficient, side ledge	$[W/(m^{2\circ}C)]$
h_{sl}	Specific enthalpy, side ledge	$[J/kg]$
I	Line current	$[kA]$
k_0	Constant in the correlation equation	$[^\circ C]$
k_1	Constant in the correlation equation	$[^\circ C/wt\%AlF_3]$
k_2	Constant in the correlation equation	$[^\circ C/wt\%Al_2O_3]$
k_{sl}	Thermal conductivity of the side ledge	$[W/(m^\circ C)]$
m_{xx}	Mass	$[kg]$
Q_{sl}	Heat flow into the side ledge	$[W]$
Q_{sw}	Heat flow through the side wall	$[W]$
R_b	Pseudo resistance	$[\mu\Omega]$
SH	Superheat	$[^\circ C]$
T_0	Ambient temperature	$[^\circ C]$
T_b	Bath temperature	$[^\circ C]$
T_{liq}	Liquidus temperature	$[^\circ C]$
U_{Cell}	Cell voltage	$[V]$
U_{ext}	A constant between 1.5-1.8	$[V]$
x_{sl}	Side ledge thickness	$[m]$
ρ_{sl}	Density of the side ledge	$[kg/m^3]$

3.2 Introduction

Controlling the alumina reducing process is challenging due to nonlinear process characteristics, coupled mass and energy balance, and few measurements.

While the control of Al_2O_3 is considered 'solved', the discussions in the literature in the last ten to fifteen years have been concerned about the control of the bath temperature and AlF_3 control. See for example Meghlaoui et al. (2002), Entner and Gudmundsson (1996), Entner (1995), Entner (1992), Wilson (1992), Peyneau (1988) and Desclaux (1987). Common for these contributions is that the AlF_3 addition is calculated as a function of

deviation from target acidity and/or target bath temperature.

It is well known in the aluminium community that both AlF_3 additions and the bath temperature have an influence on the acidity due to variation in side ledge thickness. The relationship between the bath temperature and the acidity is referred to as bath temperature–acidity correlation, or simply the correlation line. In Desclaux (1987) and Salt (1990) this is thoroughly discussed based on achieved measurements showing the correlation of bath temperature and acidity. Further investigation of the correlation line is done in Kolås (2007a).

Taylor (1992) showed that the changes in the mass balance of AlF_3 could not explain the observed changes in the acidity. According to the work done by Drengstig (1997) the variation in acidity is dominated by the variation in the bath temperature (due to side ledge variations).

The scope of this paper is to present a novel bath temperature and AlF_3 control strategy for an aluminium electrolysis cell based on estimation and decoupling techniques, together with detailed process and plant knowledge. Further it is assumed that the alumina concentration is held at a constant target value.

To enlighten the reader unfamiliar with the aluminium electrolysis process, a system description is presented in Chapter 3.3. Further the relevant theoretical background is presented in Chapter 3.4 and the results in Chapter 3.5. Finally, the results are discussed in Chapter 3.6.

3.3 System description

3.3.1 The Hall-Heroult process

The Hall-Heroult process[2], named after its inventors, is the only method by which aluminium is produced industrially today. Liquid aluminium is produced by the electrolytic reduction of alumina (Al_2O_3) dissolved in an electrolyte, referred to as bath, which mainly consists of cryolite (Na_3AlF_6). A sketch of the alumina reduction cell is shown in Figure 3.1.

In the alumina reduction cell, hereafter referred to as the cell, one (Søderberg) or several (Prebake) carbon anodes are dipped into the bath. The alumina is consumed electrochemically at the anode (Solheim (2005)),

$$\frac{1}{2}Al_2O_3 + 3NaF + \frac{3}{4}C \rightarrow AlF_3 + \frac{3}{4}CO_2 + 3Na^+ + 3e \qquad (3.1)$$

[2] The material presented regarding the Hall-Heroult process is mainly taken from Grjotheim and Kvande (1993).

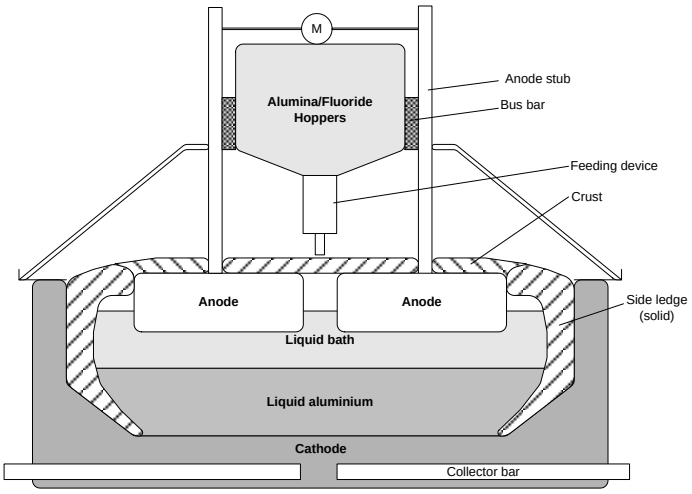

Figure 3.1: A sketch of the main features of an alumina reduction cell (Prebake).

As can be seen from (3.1), the carbon anode is consumed during the process (theoretically 333 kg C/t Al).

The lower part of the cell, the cathode, consists of a steel shell lined with refractory and thermal insulation. A pool of liquid aluminium is formed on top of the carbon bottom. The cathode, in the electrochemically sense, is the interface between the liquid aluminium and the bath,

$$AlF_3 + 3Na^+ + 3e \rightarrow Al + 3NaF \tag{3.2}$$

and the total cell reaction becomes

$$\frac{1}{2}Al_2O_3 + \frac{3}{4}C \rightarrow Al + \frac{3}{4}CO_2 \tag{3.3}$$

Pure bath (Na_3AlF_6) has a melting point of 1011 °C. To lower the melting point, the liquidus temperature, aluminium fluoride (AlF_3) and calcium fluoride (CaF_2), to mention the most important ones, are added to the bath. The bath composition in a cell may typically be 6-13 [wt%] AlF_3, 4-6 [wt%] CaF_2, and 2-4 [wt%] Al_2O_3. Lowering the liquidus temperature makes it possible to operate the cell at a lower bath temperature, but at the expense of reduced solubility of Al_2O_3 in the bath, demanding good Al_2O_3 control. It should be mentioned that if the concentration of Al_2O_3 gets too

low (less than approx. 1.8 wt%), the cell enters a state called anode effect. During anode effect, the cell voltage increases from the normal 4-4.5V up to 20-50V. Anode effect is a highly unwanted state, not only because it represents a waste of energy and a disturbance of the energy balance, but also because greenhouse gases (CF_4 and C_2F_6) are produced at the anode. Very often the anode effect requires a manual intervention of an operator.

The bath temperature during normal cell operation is between 940 °C and 970 °C.

The bath is not consumed during the electrolytic process, but some is lost, mainly during vaporization. The vapor mainly consists of NaAlF4. In addition some bath is lost by entrainment of small droplets, and water present in the alumina feed and in the air sucked into the cells reacts to form HF (Solheim (2005)). In order to protect the environment the gas is collected and cleaned in a gas scrubbing system. More than 98% (Hyland et al. (2001)) of the AlF_3 is recovered in the scrubbing system and recycled back to the cells. In addition the content of sodium oxide (Na_2O) and calcium fluoride (Ca_2F) in the fed Al_2O_3 neutralize AlF_3 (Hyland et al. (2001)). The neutralized amount (see Section 3.4) is also a function of the penetration of sodium into the cathode, and hence the cell age. As an example a 170 kA cell emits about 60 equivalent kg AlF_3 pr. 24 hours, and uses approximately 2500 kg Al_2O_3 pr. 24 hours. The amount of AlF_3 due to neutralization for a 170 kA cell is between 0 and 20 kg per 24 hour (dependent of cell age). However, since most of the AlF_3 is recycled, the real consumption of AlF_3 is very small compared to the consumption of Al_2O_3.

At the sidewalls of the cathode there is a frozen layer, called side ledge, which protects the carbon sidewall from erosion.

The composition of the side ledge is mainly pure Na_3AlF_6 with some CaF_2 (Thonstad and Rolseth (1983)). The thickness of the side ledge is a function of the heat flow through the sides, which is a function of the difference in bath temperature and liquidus temperature . Since it is assumed that the side ledge composition is mainly Na_3AlF_6, this means that the total mass of cryolite in the bath varies, while the masses of AlF_3 and Al_2O_3 do not vary with the side ledge thickness. Further, since the concentration of an additive is the mass of the additive divided by the total mass of bath, the variation in the side ledge thickness introduces variation in the concentrations. Hence, the changes in the concentrations introduce changes in the liquidus temperature, which introduces changes in the superheat, affecting the side ledge thickness.

The challenge is thereby to ensure stable cell operations resulting in a stable protective side ledge, while minimizing energy input and maximizing production.

3.3.2 Process control

In controlling a cell there are three main controlled variables: bath temperature, concentration of AlF_3 and concentration of Al_2O_3, and three control inputs: anode beam adjustments (controlling energy input), addition of AlF_3 and addition of Al_2O_3.

The dynamics in reducing the mass of AlF_3 is slow, and the control of the concentration of AlF_3 has to deal with slow responses when changing the AlF_3 concentration. The dynamics in the mass of Al_2O_3 is fast, and the control of the concentration of Al_2O_3 has to deal with quick responses. The control of the concentration of Al_2O_3 is usually considered as an isolated problem.

The bath temperature is usually measured manually once a day or at least once a week. The concentration of AlF_3 (acidity) is typically measured manually once or twice a week, while the concentration of Al_2O_3 is not normally measured at all, and only in conjunction with experiments.

The only continuous measurements is the bath pseudo resistance R_b defined as

$$R_b = \frac{U_{cell} - 1.7}{I} \qquad [\mu\Omega] \tag{3.4}$$

R_b is used as an input for the anode beam adjustment, and acts as a control variable in conjunction with the energy input to the cell.

Because the energy balance and the mass balance are coupled through the side ledge (see e.g. (Drengstig, 1997, Ch. 5)), the control of a cell must be considered as a nonlinear multivariable control problem.

3.3.3 Motivation

Given reasonable operational targets, it is believed that minimizing the process variations around target values results in good process operations in the sense of minimum pollution to the environment, maximum production and minimum expenditure. Used in the context of the alumina reduction cell the focus should be on achieving low anode effect frequency, good gas scrubbing efficiency and low deviation from target when it comes to alumina concentration, bath temperature and acidity. If the control of the alumina

concentration is reasonably good, one has to focus on the bath temperature control and the AlF_3 control.

An increase in the bath temperature results in a lower acidity and an increase in the bath conductivity (Hives et al. (1993)). According to the work done by (Drengstig, 1997, Ch. 5) the variation in acidity is dominated by the variation in the bath temperature. The control philosophy suggested is to control the bath temperature by using power additions around a target (both positive and negative) and constant addition of aluminium fluoride (AlF_3). While this seems logical and reasonable, there is a long way to go from an idea to a practical working application.

3.3.4 Experimental conditions

All the results referred to in this paper are from the development phase of the control strategy. The control strategy was developed mainly on end-to-end cells (˜170 kA) at Hydro Aluminium AS in Norway, but is implemented on side-by-side cells as well. During the development phase, the strategy was run on 10 cells with 20 comparable cells from the rest of the potline as reference cells. During the verification phase, it was run on 55 cells with 55 cells as a reference.

3.4 Theoretical background

In order to develop a control philosophy for both bath temperature and aluminium fluoride, it is essential to understand the basic process dynamics and system configuration.

3.4.1 Coupled mass & energy balance

For those familiar with aluminium electrolysis cells, it is well known that there is a relationship between the bath temperature and the acidity. In (Kolås, 2007) this relationship is investigated, and it is shown that it can be described by the equation

$$C_{AlF_3} = f_1\left(x_{sl}\right) T_b + f_2\left(x_{sl}\right) + f_3\left(C_{Al_2O_3}\right) + f_4\left(x_{sl}\right) \dot{x}_{sl} \tag{3.5}$$

where f_1, f_2 and f_4 are nonlinear functions, while f_3 is a "constant" dependent on $C_{Al_2O_3}$ (see (Kolås, 2007) for further references). An interpretation of the relationship given by (3.5) is that if the bath temperature (energy balance) changes so will the C_{AlF_3} (mass balance), and the other

way around – if the C_{AlF_3} changes, the bath temperature will change. This makes the coupling of the mass and energy balance.

By approximating (3.5) to a first order model, we get a relationship like:

$$\hat{C}_{AlF_3} = k_1 T_b + k_0 \quad [wt\%] \tag{3.6}$$

This relationship ((3.6)) states that high bath temperatures results in low acidity and low bath temperatures results in high acidity (k_1 is always negative). Further if k_0 and k_1 in (3.6) are known, we are able to estimate the C_{AlF_3} by the bath temperature T_b.

3.4.2 Aluminium fluoride (AlF_3)

Mass balance

By using first principle modelling, the rate of change of the mass of aluminium fluoride ($\dot{m}_{AlF_3}$) in an aluminium electrolysis cell can be modelled by:

$$\dot{m}_{AlF_3} = -w_{AlF_3\_out} + w_{AlF_3\_in} \quad [kg/s] \tag{3.7}$$

where $w_{AlF_3\_out}$ $[kg/s]$ is the mass leaving the cell and $w_{AlF_3\_in}$ $[kg/s]$ is the mass entering the cell. The mass of AlF_3 (m_{AlF_3}) is not measured directly, but through the relationship

$$C_{AlF_3} = 100\frac{m_{AlF_3}}{m_b} \quad [wt\%] \tag{3.8}$$

often referred to as the acidity or excess AlF_3.

While the mass entering the cell, to a large extent, is given by the mass entering through the feeders, the mass leaving the cell is a subject for closer investigation.

There are two main mechanisms for AlF_3 leaving the cell. That is AlF_3 leaving the cell by emission and AlF_3 "leaving" the cell by neutralization of AlF_3 through the addition of Al_2O_3.

Emission

The determination of the emission of AlF_3 cannot be done on AlF_3 directly, since AlF_3 leaves the cell both as particulate AlF_3 and gaseous HF (Haupin and Kvande (1993)). The total particulate consists of condensed vapor,

bath dust and alumina dust. The vapor is $NaAlF_4$ that hydrolyses partly. The bath dust contains CaF_2, Na_3AlF_6, Al_2O_3 and AlF_3, and the bath dust composition can be assumed equal to the bath composition. The alumina dust is from the fed Al_2O_3.

The gas is HF from the hydrolyzed bath and $NaAlF_4$, and the hydrolysis is determined by the reactions (Solheim (2008))

$$2AlF_3\ (in\ bath) + 3H_2O(g) = Al_2O_3 + 6HF(g) \tag{3.9}$$

$$3NaAlF_4(g) + 3H_2O(g) = Na_3AlF_6 + Al_2O_3 + 6HF(g) \tag{3.10}$$

When $NaAlF_4$ cools down, chiolite ($Na_5Al_3F_{14}$) is formed

$$5NalF_4 = Na_5Al_3F_{14} + 2AlF_3 \tag{3.11}$$

It does not matter whether HF is formed directly in the bath or by reaction with $NaAlF_4(g)$, (3.9) still represents the total turnover. Note that the basic idea behind a dry scrubber can be illustrated going from right to left in (3.9).

The equivalent total amount of emitted mass of AlF_3 is found by the relation (Haupin, Kvande 1993):

$$m_{AlF_3\_emitted} = F_{emitted}\frac{M(AlF_3)}{3M(F)}P_{Al} \quad [kg/24h] \tag{3.12}$$

where P_{Al} is produced aluminium in $[t\ Al/24h]$

$$F_{emitted} = FVP + FEP + FGP \tag{3.13}$$

FVP is the volatilized bath, FEP is entrained bath and FGP is gaseous fluorine (F) by hydrolysis of volatilized bath. The details of the emission model are to be found in Haupin and Kvande (1993).

By assuming that the pair (T_b, C_{AlF_3}) moves along its correlation line, that is, that the model described by (3.6) with $k_0 = 144$ and $k_1 = -0.14$ is valid, and assuming current efficiency (CE) equal to 93%, current (I) equal to 177.5 kA, and using Equations (3.12)-(3.13), the estimated equivalent emitted mass AlF_3 from a cell is as given by Table 3.1.

The results in Table 3.1 indicate that the variation in emission from a cell is small when the pair (T_b, C_{AlF_3}) moves along the correlation line.

Table 3.1: Estimated emitted AlF3 at CE=93,I=177.5kA.

Bath temp $[^oC]$	Correlated acidity $[wt\%]$	Emitted AlF_3 [kg/24h]
930	16.1	65.1
940	14.7	65.6
950	13.3	65.9
960	11.9	66.0
970	10.6	65.9
980	9.2	65.7
990	7.8	65.4

Neutralization

Another source of the variation in the mass of AlF_3 in the bath is related to the Na_2O and CaO in the added alumina (Hyland et al. (2001)). The reason is that Na_2O and CaO in the alumina neutralizes AlF_3 in the bath. There are two main reactions to take into account:

$$3Na_2O + 4AlF_3 = 2Na_3AlF_6 + Al_2O_3 \tag{3.14}$$

$$3CaO + 2AlF_3 = 3CaF_2 + Al_2O_3 \tag{3.15}$$

The reaction described by (3.14) produces neutral bath that has to be enriched with AlF_3 to achieve the desired acidity (Hyland et al. (2001)). That is

$$m_{Na_3AlF_6\_produced} = m_{Al_2O_3}\frac{[Na_2O]}{100}\frac{2M_{Na_3AlF_6}}{3M_{Na_2O}} \tag{3.16}$$

$$m_{Na_3AlF_6\_produced} + m_{AlF_3\_extra} = m_{b\_with\_ref.acidity} \tag{3.17}$$

A calculation on the system described by Equations (3.14)-(3.17) gives the neutralized amount of AlF_3 described by

$$m_{AlF_3\_neutralized} = m_{Al_2O_3}\left\{\begin{array}{c}\frac{[Na_2O]}{100}\frac{4M_{AlF_3}}{3M_{Na_2O}} + \frac{[CaO]}{100}\frac{2M_{AlF_3}}{3M_{CaO}} + \\ \frac{[Na_2O]}{100}\frac{2M_{Na_3AlF_6}}{3M_{Na_2O}}\frac{C_{AlF_3\_ref}}{1-C_{AlF_3\_ref}}\end{array}\right\} \tag{3.18}$$

Table 3.2: Estimated neutralized mass of AlF3 at CE=93 percent, I=177.5 kA, Ref. acidity=11.9 wt percent for different alumina batches.

Na_2O [wt%]	CaO [wt%]	Neutralized AlF_3 [kg/24h]	Alumina batch
0.37	0.004	21.4	Batch 1
0.43	0.031	25.6	Batch 2
0.46	0.029	27.3	Batch 3
0.42	0.034	25.1	Batch 4
0.39	0.005	22.6	Batch 5
0.42	0.006	24.3	Batch 6
0.43	0.031	25.6	Batch 7
0.37	0.030	22.1	Batch 8
0.40	0.031	23.9	Batch 9
0.40	0.042	24.1	Batch 10
0.35	0.010	20.4	Batch 11

There are two main problems related to the neutralization. First, the neutralization of AlF_3 is dependent on the added mass of Al_2O_3. Second, the neutralization of AlF_3 is dependent on the alumina quality. In this work, we assume that the variation in the added mass of Al_2O_3 can be neglected.

In Table 3.2, the alumina specifications from different alumina vendors are compared, and the theoretical amount of neutralized AlF_3 is estimated by using (3.18).

As Table 3.2 shows, the neutralized mass of AlF_3 varies considerably with the alumina batch. If, in the worst case, it should happen to be a change from the minimum neutralizing Al_2O_3 (Batch 11) to the maximum neutralizing Al_2O_3 (Batch 3) the cell would be 200 kg short in AlF_3 after one month, assumed constant addition.

The picture of which parameters that have influence on the variation of the mass of AlF_3 in a cell is completed by taking into account the cell age (see e.g. Hyland et al. (2001)), because the penetration of the Na into the cathode decreases with increasing cell age.

Table 3.3 shows the added AlF_3 in percent of estimated AlF_3 ((3.18)) grouped by cell age in a production line at Hydro Aluminium AS.

As can be seen from Table 3.3, the actual addition is close to the estimated AlF_3 addition.

Table 3.3: AlF3 addition [kg/24h] grouped by cell age.

Cell age [days]	No. of cells	$(AlF_3added/AlF_3estimated) * 100[\%]$
450-499	28	97
500-1499	36	96
> 1500	35	86

3.4.3 Bath temperature

The cell voltage is the control property for the heat-up of the bath and it is the driving force in the reaction in the cell. The cell voltage can be described by:

$$U_{Cell} = \sum_i |U_{Polarization,i}| + \sum_j U_{Ohmic\ loss,j} + \sum_k U_{External\ loss,k} \quad (3.19)$$

that is:

$$U_{Cell} = |E^{rev}| + |n_{cc}| + |n_{aa}| + |n_{ac}| + U_b + U_e \quad (3.20)$$

where

U_{cell}	cell voltage
E_{rev}	reversible (equilibrium) voltage
n_{cc}	concentration overvoltage at the cathode
n_{aa}	reaction overvoltage at the anode
n_{ac}	concentration overvoltage at the anode
U_b	ohmic voltage drop in the electrolyte
U_e	sum of cathode, anode and other external voltage drops.

In the literature there are many versions of the components contributing to the cell voltage. For simplicity the following model is as described in Grjotheim and Kvande (1993). The model is not the most up to date type, but it has the dynamics that are important for this work, and it suits the purpose of illustrating the process behavior.

The E^{rev} and the various overvoltages are modelled as semi-empirical formulas:

$$|E^{rev}| = -\frac{\Delta G^o}{6F} + \frac{RT_{b_{abs}}}{6F} \ln\left(\frac{C_{Al_2O_3(sat)}}{C_{Al_2O_3}}\right)^{2.77} \quad (3.21)$$

$$|n_{cc}| = \frac{RT_{b_{abs}}}{1.5F}\left(1.375 - 0.125\ CR\right)\ln\left(\frac{i_c}{0.257}\right) \tag{3.22}$$

$$|n_{aa}| = \frac{RT_{b_{abs}}}{1.08F}\ln\left(\frac{i_a}{i_0}\right) \tag{3.23}$$

$$|n_{ac}| = \frac{RT_{b_{abs}}}{2F}\ln\left(\frac{i_{cc}}{i_{cc} - i_a}\right) \tag{3.24}$$

where

ΔG^o	standard Gibbs energy for the reaction
$C_{Al_2O_3}$	alumina concentration [wt%]
$C_{Al_2O_3(sat)}$	alumina saturation concentration [wt%]
$T_{b_{abs}}$	electrolyte (bath) temperature [K]
CR	cryolite ratio (mol)
F	Faraday constant
i_a	anodic current density $[A \cdot cm^{-2}]$
i_c	cathodic current density $[A \cdot cm^{-2}]$
i_{cc}	critical current density $[A \cdot cm^{-2}]$
i_o	exchange current density $[A \cdot cm^{-2}]$
R	universal gas constant

The critical current density (i_{cc}) and the exchange current density (i_o) can be modelled as:

$$i_{cc} = \left[5.5 + 0.18\left(T_{b_{abs}} - 1323\right)\right]\left(C_{Al_2O_3}^{0.5} - 0.4\right)A_a^{-0.1} \tag{3.25}$$

$$i_0 = 0.0029 C_{Al_2O_3}^{0.56} \tag{3.26}$$

where

Aa anode surface $[cm^2]$.

In order to take into account the effect of the gas bubble layer, the ohmic voltage drop, U_b, caused by the bubbles, can be modelled as (Haupin (1987)):

$$U_b = \frac{i_a}{\kappa}\left(ACD - \delta\right) + U_{bu} \tag{3.27}$$

$$U_{bu} = \frac{i_a}{\kappa}\left[(\delta - t_a)\,(1-\varepsilon)^{-1.5} + t_a\,(1 - 1.26\ f_c)^{-1}\right] \tag{3.28}$$

$$\kappa = e^{\left(2.0156 - 0.0207 C_{Al_2O_3} - 0.005 C_{CaF_2} + 0.2175 CR - \frac{2068.4}{T_{b_{abs}}}\right)} \tag{3.29}$$

$$fc = \frac{1}{1 + 0.75 C_{Al_2O_3}} \tag{3.30}$$

where

ACD	Anode Cathode Distance
κ	electrical conductivity of the bath (Choudhary's formula)
U_{bu}	extra voltage drop due to the bubbles
δ	bubble layer thickness ($= 1.5\ [cm]$)
t_a	adhering bubble thickness ($= 0.5[cm]$)
ε	gas fraction in the bath ($= 0.02 C_{Al_2O_3}$)
f_c	fraction of anode covered with gas
C_{CaF_2}	$[wt\%]\ CaF_2$

Finally, the (pseudo) bath resistance is modelled as

$$R_b = \frac{U_{cell} - 1.7}{I} \tag{3.31}$$

where

Rb	(pseudo) bath resistance $[\mu\Omega]$
I	line current $[A]$

By using the model Equations (3.19)-(3.30), one is able to investigate the relationship between the bath resistance, bath temperature, acidity and alumina concentration (Figure 3.2) at constant ACD and assuming (T_b, C_{AlF_3}) moves along the correlation line.

As Figure 3.2 illustrates, the bath resistance decreases with increasing bath temperature and decreasing acidity.

3.4.4 Measuring the performance

As stated in the introduction, it is desirable to achieve minimum deviation from target values. A method for measuring the performance of a cell is to

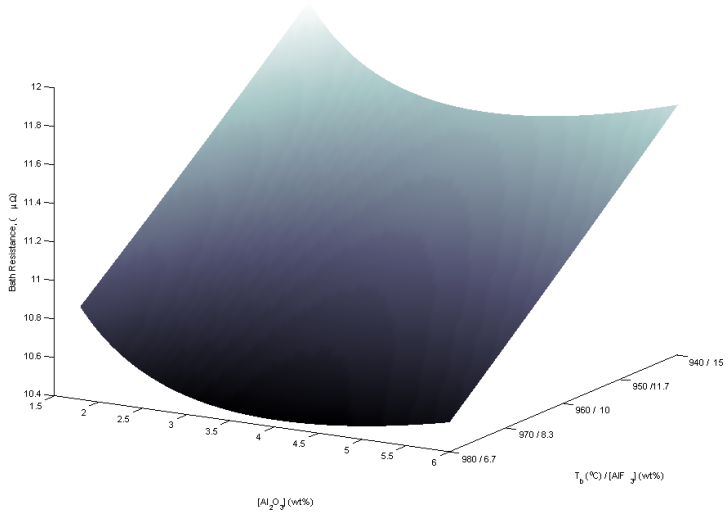

Figure 3.2: The relationship between bath resistance (R_b), bath temperature (T_b), acidity (C_{AlF_3})and alumina concentration ($C_{Al_2O_3}$) at constant ACD.

use the formula for standard deviation. It is well known in the aluminium community that cells that are outside their operating window have little variations in temperature. E.g. a hot cell might have a very stable bath temperature. In a performance-measuring context this would lead to a too good result. Hence, the mean value in the standard formula is replaced by the target value.

$$\tilde{\sigma}_{T_b} = \sqrt{\frac{\sum_{i=1}^{n} \left(T_b - T_{b\_target}\right)^2}{n-1}} = \frac{\left\|T_b - T_{b\_target}\right\|_2}{\sqrt{n-1}} \quad [^oC] \tag{3.32}$$

It can be shown that this formula is equal to the normal formula for standard deviation when the target value is equal to the true mean value. By using this technique of measuring the performance of a cell, one addresses both minimum variation and target value.

3.5 Results

3.5.1 Coupled mass & energy balance

A typical general equation for AlF_3 addition is (Hyland et al. (2001)):

$$u_{AlF_3} = \alpha + \beta \left(T_b - T_{b\_target}\right) + \lambda \left(C_{AlF_3} - C_{AlF_3\_target}\right) \tag{3.33}$$

where α is some constant base feed, and β and λ are dependent of the technology implemented. A drawback with these types of controllers is that they focus on the target acidity instead of the target m_{AlF_3}, and hence do not take into account the coupling between the mass and the energy balance in the cell. For example, assume a cell with correct amount of AlF_3, and that is low on acidity due to high bath temperature (which causes the ledge to melt and hence high mass of bath and therefore low measured acidity). Most of these controllers add AlF_3 to achieve the target acidity. When the cell is back on the target temperature, the cell will be too acid due to the now too large content of AlF_3.

Drengstig et al. (1998) suggested maintaining the mass balance of AlF_3 rather than the acidity. To implement this idea one has to be able to identify if the cell has too much or too little mass of AlF_3.

A possible method to identify if the cell has too much or too little mass of AlF_3 is to use the relationship between the bath temperature and

acidity. A hypothesis could be defined that if a cell is below its correlation line (given by (3.6)) it has too little mass of AlF_3, and if it is above its correlation line it has too much mass of AlF_3. And finally, if it is on the correlation line, it has the correct amount of mass of AlF_3. A picture of this situation is given in Figure 3.3.

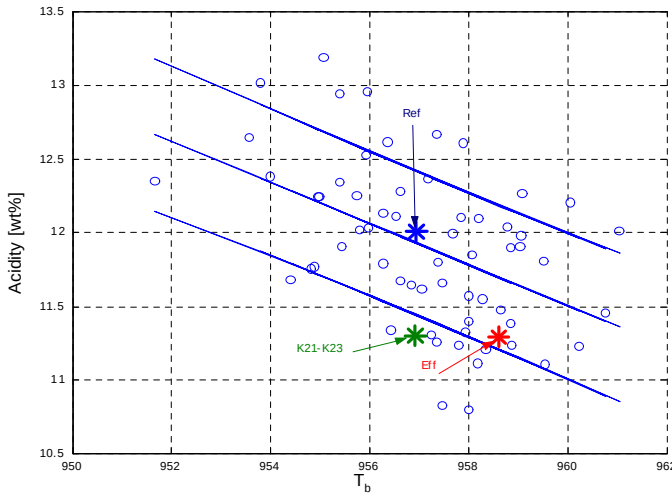

Figure 3.3: Correlation curve between the bath temperature and the acidity. The line in the middle describes the correlation line for this technology. The upper star is the target value, and the lower left and lower right stars indicate the location of the test cells. Clearly the test cells are below their target values.

What this hypothesis does is to use the relationship between the mass balance and the energy balance to estimate the acidity based on the measured bath temperature. In addition, the estimator acts as a decoupler between the energy balance and the mass balance. By comparing the estimated acidity with the measured acidity one achieves control of the mass of AlF_3 rather than a target acidity.

This hypothesis was tested on a number of cells at Hydro Aluminium AS, and a result of one of these tests is shown in Figure 3.4.

By adding a calculated extra amount of AlF_3, the cell responds by establishing itself (from the fourth data point shown) on the desired correlation line (upper line). Still the cell has not achieved its temperature target, and

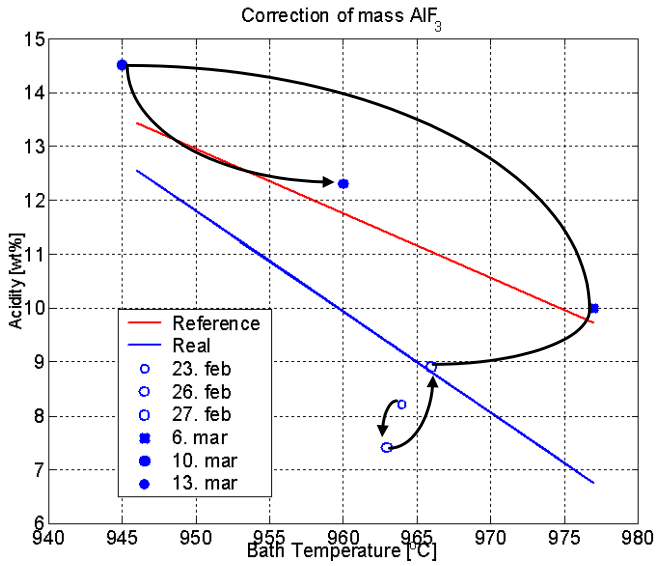

Figure 3.4: The response of an AlF_3 addition to an identified cell with too low amount of AlF_3. The cell was below its correlation line (lower line) and also below the correlation line typical for that production line (upper line), which is the desired correlation line for this technology. The open circles are before the addition; the filled circles are after the addition.

no additional AlF_3 is added but the estimated base feed. When temperature control is satisfactory, the acidity is stabilized at the approximate target value.

3.5.2 AlF_3 control

If we assume that all the cells are fed by secondary alumina from a closed gas scrubbing system, and that the scrubbing system has high regularity, more than 98% of what goes off the cell is returned to the cell via the scrubbing system (Hyland et al. (2001)). Since most of the AlF_3 that leaves the cell comes back into the cell via the secondary alumina (from the scrubbing system) and the variation in emission is small (Table 3.1), we can in the long run "forget" the emission from the cell when it comes to AlF_3 control.

If we assume negligible variations in fed Al_2O_3 and that the Na_2O and the CaO content of the secondary alumina is equal to that in the primary alumina, the main reason for the variation in the mass of AlF_3 in the bath is related to the Na_2O and CaO variations in the primary alumina (Hyland et al. (2001)).

Hence the primary task for an AlF_3 controller is to compensate for the neutralized amount of AlF_3, together with securing the correct mass of AlF_3 in the cell. Further, the AlF_3 controller could utilize plant knowledge in a feed forward strategy to compensate for changes in the alumina quality (see (3.36) below and Table 3.2).

A model describing the cell given the assumption above is:

Process model:

$$\dot{\hat{m}}_{AlF_3} = -\hat{w}_{AlF_3\_out} + \hat{w}_{AlF_3\_in} \quad [kg/s] \tag{3.34}$$

Estimator/ decoupler[3] :

$$\hat{C}_{AlF_3} = k_1 T_b + k_0 \quad [wt\%] \tag{3.35}$$

$$\hat{w}_{AlF_3\_out} = f(y_{CE}, y_I, y_{Na_2O}, y_{CaO}, y_{Age}) \tag{3.36}$$

Measurement vector (y):

$$y = \begin{bmatrix} T_b & C_{AlF_3} \end{bmatrix} \tag{3.37}$$

[3]Equal to Equation (3.6), but repeated for convenience

The theory stated above enables us to estimate the amount of AlF_3 needed. Drengstig et al. (1998) suggest that the AlF_3 addition should be kept constant. A problem with a constant addition of AlF_3 is the integration effect, in that if the addition is a bit too low or a bit too high, the mass of AlF_3 will diverge one way or the other, making it problematic to achieve target values. Furthermore, one has no guarantee that the mass of AlF_3 in the cell was initially estimated correctly anyway. Assuming that the constant addition was correct, if the cell were short of mass of AlF_3 it would remain too low.

By designing a controller like

$$u_{AlF_3} = w_{AlF_3\_in} = \hat{w}_{AlF_3\_out} + \Delta u_{AlF_3} \quad [kg/s] \tag{3.38}$$

where Δu_{AlF_3} is the output from a PID-controller, we are able to achieve control of the AlF_3 mass balance. The purpose of the PID controller is to compensate for model errors and initially erroneous mass of AlF_3 in the cell.

Figure 3.5 illustrates the architecture of the AlF_3 controller.

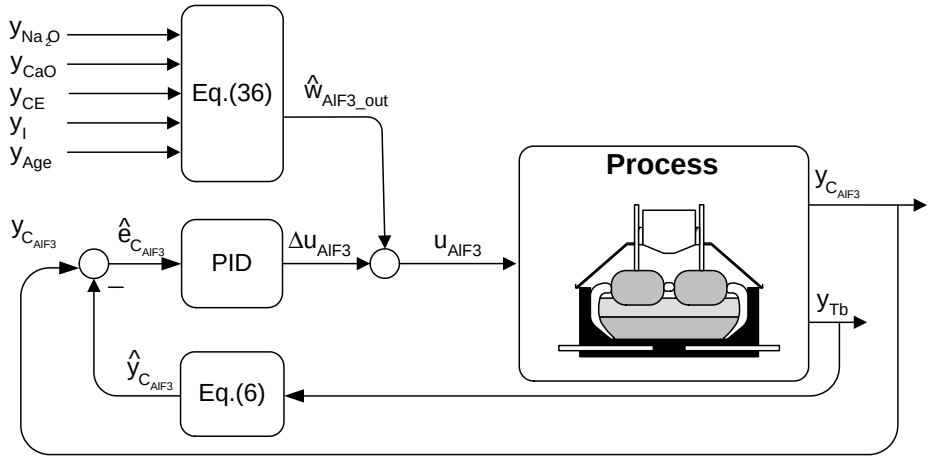

Figure 3.5: The architecture of the AlF_3 controller.

The controller was realized by using the theory above for estimating the constant base feed ($\hat{w}_{AlF_3\_out}$) as stated by (3.36). Note that (3.36) would be technology specific, and is mainly determined by (3.18), which need to be modified according to technology specific cell age properties due to penetration of Na in the cathode. Having control on the quality of

the alumina entering the cell, that is the Na_2O and CaO contents of the primary alumina, feed-forward control of the base feed is achieved by using this information into (3.36) when the alumina quality changes.

To calculate the extra feed, a PID controller with anti-windup was used, acting on the deviation between the measured acidity and the estimated acidity, where the estimated acidity was estimated from (3.6). By this one achieve a decoupling of the mass and energy balance with respect to controlling the mass of AlF_3 in the cell.

In addition, to dampen the influence of variations introduced with operational practice (measurement noise), the bath temperature and acidity measurements were low pass filtered before used in the controller.

3.5.3 Bath temperature control

Figure 3.2 illustrates the effect of increasing bath conductivity with increasing bath temperature. In a cell where the control philosophy is to maintain a target bath resistance, the result is that the ACD increases with increasing bath temperature (and decreasing acidity). This is because increasing bath temperature results in a decreased resistance, but the energy controller tries to maintain target resistance, which is achieved by increasing the anode-cathode distance (ACD).

By using the model Equations (3.20)-(3.31) on real plant data, and assuming constant alumina concentration, the estimated ACD as a function of temperature is as shown in Figure 3.6.

Drengstig (1997), Drengstig et al. (1998) and Drengstig et al. (2002) claims that the energy input to the cell could be used to control the bath temperature. As shown in the figure above, because the ACD is high when the bath temperature is high, there is room for decreasing the energy input, and hence the bath temperature, by lowering the ACD.

By accepting the hypothesis that hot cells have high ACD and cold cells have low ACD, a logical energy control scheme would be to decrease the ACD if the cell is too hot, and vice versa.

The energy controller is illustrated in Figure 3.7.

By using the deviation from the target bath temperature (e_{T_b}) to drive a PID controller to calculate an addition (ΔRb) to the reference pseudo resistance (Rb_{ref0}), it is possible to achieve reasonable good bath temperature control.

To dampen the influence of variations introduced with operational practice (measurement noise) in the bath temperature measurements, the bath temperature is low pass filtered.

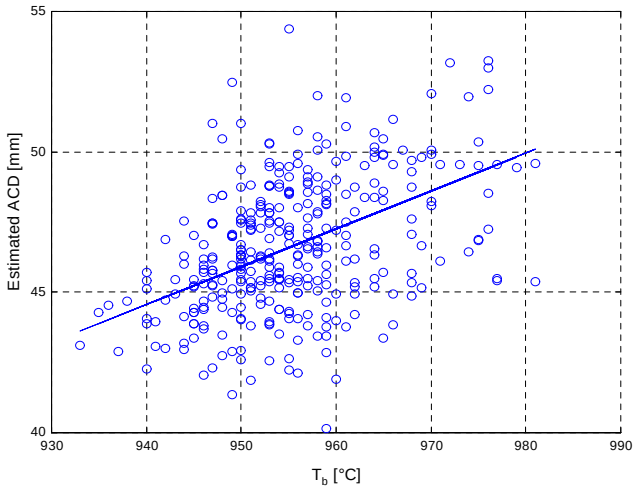

Figure 3.6: Estimated ACD for a cell as a function of bath temperature. The figure indicates that for this particular cell, the increase in the ACD is estimated to approx. 8 mm when the bath temperature is varying from 940 ^{o}C to 980 ^{o}C. The estimate is based on measured bath resistance, bath temperature, acidity and assumed constant alumina concentration (3 wt%).

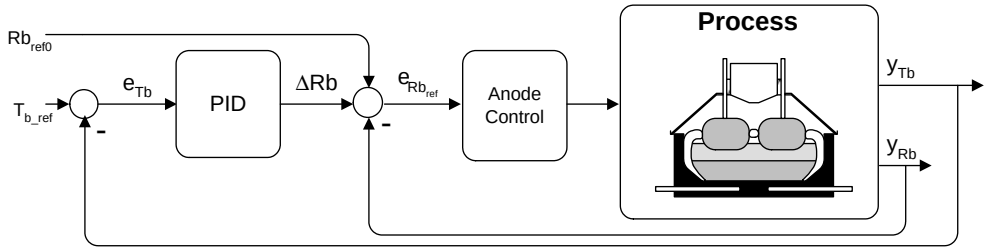

Figure 3.7: Architecture of the bath temperature controller.

Table 3.4: Achieved performance

Period	$\tilde{\sigma}_{T_b}$ test cells $[^\circ C]$	$\tilde{\sigma}_{T_b}$ reference cells $[^\circ C]$
Jan-Oct	7.2	8.1
Jun-Oct	6.4	8.2

Performance

By using the proposed control philosophy on end-to-end cells[4] , the results with respect to $\tilde{\sigma}_{T_b}$is shown in Table 3.4. The operators measured the bath temperature on each cell two times a week.

The results are calculated by first calculating $\tilde{\sigma}_{T_b}$ for each cell based on all the measurements that month, then making an average $\tilde{\sigma}_{T_b}$ for that month based on all these $\tilde{\sigma}_{T_b}$'s, and finally calculating the average $\tilde{\sigma}_{T_b}$ for the test period. For example the period Jun-Oct is the average of the 5 monthly average $\tilde{\sigma}_{T_b}$ achieved.

The period Jan-Oct includes both the period of implementing and tuning the strategy. The period Jun-Oct includes only the final tuned strategy.

When it comes to energy consumption, the tested strategy achieved approximately 2.2% reduction in cell voltage.

3.6 Discussion and conclusions

This paper presents a novel control strategy for controlling both the bath temperature and the mass of AlF_3 in an aluminium electrolysis cell. The control philosophy is based on both detailed process and plant knowledge.

The main difference from existing strategies that focus on acidity is that this strategy focuses on achieving constant mass of AlF_3 in the cell by using an estimator and a decoupler, and uses energy to control the bath temperature. In addition a feed-forward strategy is used to react early on raw material changes.

While the control strategy seems simple and logical, it is challenging when it comes to tuning. Since the process is non-linear the controlled process gets easily unstable if not tuned carefully.

However, the results show that by using this control strategy, it is possible to reduce the variations in the bath temperature with approximately 20% and thereby operating closer to the target values, resulting in lower power consumption without loss of current efficiency (CE).

[4]See Section 4.3.4.

3.7 Acknowledgements

The authors acknowledge the following persons for the support in developing, implementing and verifying this strategy: V.G. Christensen, R. Hovland, E. Rinde, G. Valsvik, H. Gikling, E. Svinsås, F. Øvstetun, M.Liane, P.A. Solli, T. Moen, J.Aalbu and all the various project members through times, all connected to Hydro Aluminium AS. We also acknowledge Professor B.A. Foss at NTNU and H. Kvande at Hydro Aluminium AS for the discussions and feedback during the writing of this paper.

Part II

Contributions regarding state estimation

Chapter 4

State estimation IS the real challenge in NMPC

This chapter is an extended version of an accepted abstract by Kolås, Foss and Schei 2008 for the workshop NMPC'08, Pavia, Italy

4.1 Abstract

Nonlinear model predictive control (NMPC) opens for the use of MPC in more demanding applications than has normally been the case for linear MPC. Hence, NMPC is gaining wider acceptance as a technology for challenging control problems within the process industries.

The scope of this paper is to discuss an important challenge when applying NMPC, namely the state estimation problem. Our experience shows that this is the critical challenge in most cases. To ensure a sound level of credibility we first present some applications and typical control challenges in some detail. Thereafter, we discuss approaches to get to grips with the state estimation problem in a sound manner focussing on two issues: state estimation methods and noise modeling.

The importance of applying higher order filters as an alternative to the extended Kalman Filter (EKF) in demanding applications is highligthed. To overcome limitations in the EKF, several approaches and alternatives have been suggested. We will focus on some of the differential-free algorithms - the Divided Difference (or Central Difference) Kalman filter (DDKF/CDKF) and the Unscented Kalman filter (UKF), and our experience with these approaches.

Recursive state estimation algorithms usually assumes that uncertainty

enters through additive white noise sources. We discuss noise modeling based on a hypothesis that it is important to model noise correctly. In practice this implies a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty.

Keywords: Unscented, Divided Difference, Kalman filter, Nonlinear state estimation, Constrained nonlinear systems, Noise modeling.

4.2 Introduction

Nonlinear model predictive control (NMPC) opens for the use of MPC in more demanding applications than has normally been the case for linear MPC. In particular NMPC lends itself to nonlinear systems which exhibit large variations in operating conditions and which are critically dependent on the use of a dynamic nonlinear model to gain sufficient performance. Hence, NMPC is gaining wider acceptance as a technology for challenging control problems within the process industries. NMPC is not a well defined term in the sense that NMPC may be used for controllers ranging from a slight variation of linear MPC to the online solution of a constrained nonlinear optimization problem. One example of a slight modification to account for nonlinearity is the use of multiple linear models in such a way that the current working point defines which model should be active at a given time instant. Hence, the QP-problem frequently encountered in linear MPC will change as the active model changes. In our context NMPC shall mean the use of a nonlinear mechanistic model, state estimation, and the solution of an online constrained nonlinear optimization problem.

The scope of this paper is to discuss an important challenge when applying NMPC, namely the state estimation problem. Our experience shows that this is the critical challenge in most cases. To ensure a sound level of credibility we first present some applications and typical control challenges in some detail. Thereafter, we discuss approaches to get to grips with the state estimation problem in a sound manner focussing on two issues: state estimation methods and noise modeling. The importance of applying higher order filters as an alternative to the extended Kalman Filter (EKF) in demanding applications is highligthed. To overcome limitations in the EKF, several approaches and alternatives have been suggested. We will focus on some of the differential-free algorithms - the Divided Difference (or Central Difference) Kalman filter (DDKF/CDKF) and the Unscented Kalman filter (UKF), and our experience with these approaches.

Recursive state estimation algorithms usually assumes that uncertainty enters through additive white noise sources. Further, unknown and time-varying parameters are often modelled similarly by augmenting the states with a parameter vector. Finally, initial model uncertainty is reflected through the choice of the initial covariance matrices for the states and parameters. We discuss noise modeling based on a hypothesis that it is important to model noise correctly. In practice this implies a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty.

The paper starts with a motivating example before some estimation algorithms and noise models are explored. A discussion and some conclusions end the paper.

4.3 A motivating example - the Hall-Heroult process

The Hall-Heroult process is the dominating process for producing aluminum today (Grjotheim and Kvande (1993)). The fundamentals of the process are to dissolve Al_2O_3 in molten cryolite (also known as electrolyte or bath), and electrically reduce complex aluminum containing ions to pure aluminum. The process has strong internal couplings, for instance between the mass and energy balance through the side ledge. The coupled mass and energy balance combined with nonlinear process characteristics and few measurements, makes the Hall-Heroult process challenging to control (Foss and Schei (2005), Drengstig et al. (1998), Gran (1980)).

Recently Hydro Aluminium AS have been active in developing an advanced control structure, by initiating an NMPC project that has resulted in a patent application for NMPC control of the Hall-Heroult process (Kolås (2007b)). An important challenge in an NMPC application is connected to the estimator, in that the complexity and efficiency of the NMPC is closely related to the quality of the estimates produced by the estimator. This is illustrated in Figure 4.1, where data from one of the early tests of NMPC in closed loop control of the Hall-Heroult process is shown.

Figure 4.1 clearly illustrates that the performance of the estimator is crucial for the expected performance of the NMPC application. The quality of the estimates may not only be dependent on the accuracy of the model, but also of the estimating method selected and how process knowledge is applied (Kolås et al. (2008b)).

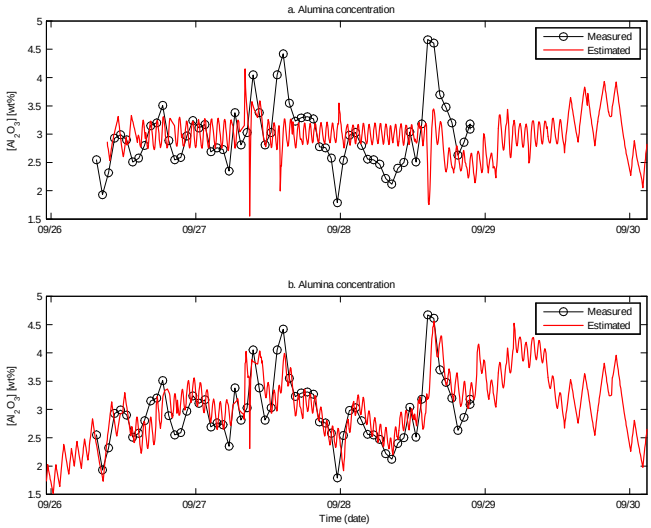

Figure 4.1: The figure shows measured and estimated alumina concentration for different tuning (a) and (b) of an estimator for the Hall-Heroult process. Note that the measured alumina concentration is not available to the estimator.

4.4 Nonlinear state estimation

Nonlinear state estimation is a field with broad contributions[1]. It includes nonlinear estimators such as Moving Horizon Estimation, the Particle filter, the Ensemble Kalman filter, the Unscented Kalman filter and the Extended Kalman filter, just to mention some. The Extended Kalman filter (EKF), which was originally proposed by Stanley Schmidth in 1967 (Bellatoni and Dodge (1967)) in order to apply the Kalman filter to nonlinear spacecraft navigation problems, is probably the most used method in applied nonlinear state estimation. However, several authors have experienced shortcomings applying EKF to systems with severe nonlinearity and/or constraints (see e.g. Chen et al. (2006), Vachhani et al. (2006), Julier and Uhlmann (1994), Julier et al. (1995), Kandepu et al. (2008), Nørgaard et al. (2000), Rao (2000), Hasseltine and Rawlings (2003), Bizup and Brown (2003) to mention some). The shortcomings are related to difficulties in determining the Jacobians, errors introduced by linearization and/or to deal with systems with multimodal or asymmetric probability density functions (pdf). Also, if handling of constraints are unavoidable, the EKF has some limitations in propagating the constraints both through the states and covariance calculations.

In this work we use the reformulated UKF algorithms as described in Kolås et al. (2008a). When using UKF with constraint, we constrain the sigma points and the updated sigmapoints. The constraint in this work is implemented by projection similar to 'clipping' (for further references on constraints and UKF see Kolås et al. (2008a)).

Also the Divided Difference Kalman filter (DDKF) is briefly considered in this work (Nørgaard et al. (2000)). Further, the analysis on how noise enters the covariance is done using a continuos time formulation, and we assume these results also are valid for the discrete time case. Based on this assumption, and for the case of simplicity regarding the theoretical results presented, we investigate the covariance equation found in the continuos EKF (Gelb (1974))

$$\dot{P} = FP + PF^T + HQH^T - KRK^T \tag{4.1}$$

where all the elements in (4.1) may be time varying. Here K is the Kalman gain typically found in he Kalman filter, F and H denote the Jacobians found from an appropriate system model. Especially the term HQH^T in

[1] For a review of nonlinear state estimation see Daum (1986) and Daum (2005).

(4.1) will have our attention in this work, since this is the term that determines how the system noise is injected into the covariance calculation. A more detailed discussion about noise and modeling is to be found in Kolås et al. (2008b). Further, we have the assumed process noise covariance matrix $E\left\{v(t)v^T(\tau)\right\} = Q(t)\delta(t-\tau)$ and the assumed measurement noise covariance matrix $E\left\{w(t)w^T(\tau)\right\} = R(t)\delta(t-\tau)$, where $\delta(t-\tau)$ is the Dirac function. Further we have the relationship between the continuos time spectral density Q and the discrete-time covariance Q_k by

$$Q = Q_k/\Delta t \tag{4.2}$$

and R and the discrete measurement covariance matrix R_k by

$$R = R_k\Delta t \tag{4.3}$$

where Δt denotes the discrete time step (Simon, 2006, pp. 233).

4.5 Estimation in constrained nonlinear systems

4.5.1 Investigated case - 2 state CSTR

The results from the 'motivating example' motivated further studies of the UKF approach. In Kolås et al. (2008a) a broad overview of different UKF algorithms is given, and extensions to the ensemble of UKF algorithms are suggested, discussing in particular the use of constraints. In the sections to come, the performance of the constrained approach is compared with EKF, a selection of UKF algorithms and some DDKF algorithms applied to a 2 state nonlinear CSTR example process with a multimodal probability density functions. This is a very challenging state estimation problem.

Case description

Consider the gas-phase, reversible reaction Hasseltine and Rawlings (2003)

$$2A \xrightarrow{k_r} B \qquad k_r = 0.16 \tag{4.4}$$

with stoichiometric matrix

$$s = \begin{bmatrix} s_1 & s_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \end{bmatrix} \tag{4.5}$$

and reaction rate

$$r = k_r P_A^2 \tag{4.6}$$

The state and output vectors are defined as

$$x = \begin{bmatrix} P_A \\ P_B \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad y = \begin{bmatrix} 1 & 1 \end{bmatrix} x \tag{4.7}$$

where P_A and P_B are the partial pressures. It is assumed that the ideal gas law holds (high temperature, low pressure), and that the reaction occurs in a well-mixed, isothermal batch reactor.

The model used is the discrete analytical solution as found in Hasseltine and Rawlings (2003). Further it is assumed that the system experiences Gaussian noise both in the states and in the outputs, respectively $v_k \sim N(0, Q_k)$ and $w_k \sim N(0, R_k)$.

The parameters used for the discrete representation of this system are $\Delta t = t_{k+1} - t_k = 0.1$, $P_0 = diag^2 \begin{bmatrix} 6^2 & 6^2 \end{bmatrix}$, $k_r = 0.16$, $x_0 = \begin{bmatrix} 3 & 1 \end{bmatrix}^T$, $x_0 = \begin{bmatrix} 3 & 1 \end{bmatrix}^T$, $\hat{x}_0 = \begin{bmatrix} 0.1 & 4.5 \end{bmatrix}^T$, $Q_k = diag \begin{bmatrix} 0.001^2 & 0.001^2 \end{bmatrix}$ and $R_k = 0.1^2$. Note that the initial guess for the estimator ($\hat{x}_0$), is very poor. This simple example is used by several authors in order to investigate estimator performance (see Hasseltine and Rawlings (2003), Vachhani et al. (2006), Rawlings and Bakshi (2006), Kandepu et al. (2008)). The reason why this problem is interesting is that the estimator may experience a multimodal probability density function, which may lead to unphysical estimates.

4.5.2 Simulation results

In the following chapters we will investigate some estimation algorithms applied to the 2-state CSTR case. Note that all the parameters are as described in the case description above for all the presented algorithms. In the presented results also the noise sequences are identical in all simulations, except for the DDKF. We have chosen to be true to the source of these examples Hasseltine and Rawlings (2003) and have used the same parameters to achieve comparable results.

EKF

The EKF algorithm with numerically derived Jacobians is investigated. Figure 4.2 shows the results of the simulation using unconstrained EKF.

[2] $diag()$ is an operator creating an nxn matrix with the given elements on the diagonal and 0 in all other entries.

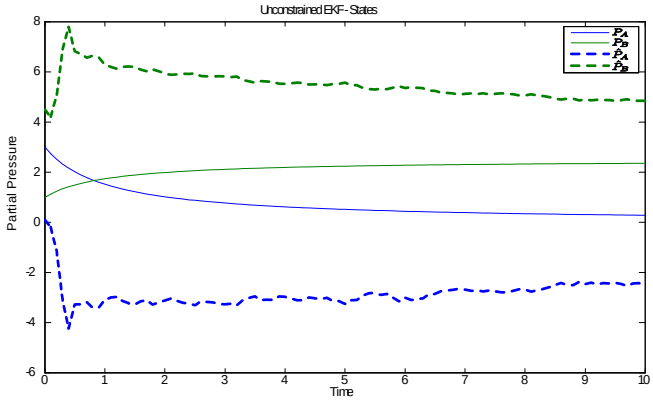

Figure 4.2: Unconstrained EKF.

The unconstrained EKF fails to converge to the true states within the given time frame[3]. These results are in agreement with the results of Hasseltine and Rawlings (2003) and Kandepu et al. (2008). The reason why the EKF fails is that while the negative pressure is unphysical, the unconstrained estimator allows the estimate to enter regions where the partial pressure may be negative. By using EKF with constraints, where the constraints are implemented by clipping the corrected state estimate such that $\hat{x}_k \geq 0$, again the constrained EKF fails in converging to the true states. These results are in agreement with the results of Hasseltine and Rawlings (2003) and Kandepu et al. (2008). By clipping the state $\hat{x}_k$, it is restricted to a valid physical region, but the knowledge about the constraints is not propagated into the covariance, and hence the accuracy of the approximated covariance matrix P_{x_k} is questionable.

UKF

The unconstrained UKF and the reformulated UKF, with the same tuning as the EKF, is applied to the test case.

As Figure 4.3a shows, the unconstrained UKF fails to converge to the true states within the given time frame. These results are in agreement with the results of Kandepu et al. (2008). The unconstrained UKF as well as the EKF has to deal with multiple optima, and we believe this is the reason

[3]Actually the EKF will converge very slovly, but one need to run the simulation approx. 1000 samples.

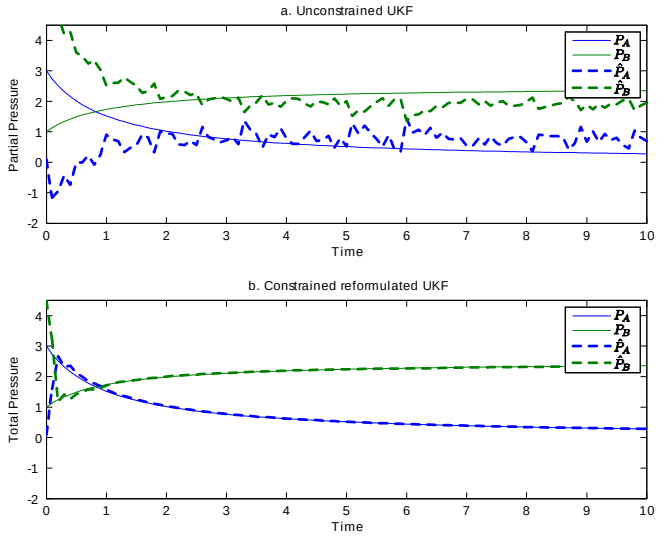

Figure 4.3: The figure shows a. the results of using the unconstrained fully augmented UKF as by van der Merwe (2004) and b. the results of using the constrained reformulated UKF as proposed in Kolås et al. (2008a). Constraints used is CC_1 and CC_7 and the non-augmented UKF.

why also the unconstrained UKF algorithms suffer poor performance on this case. All the investigated UKF algorithms converge to the true state when constraining the sigma points $\chi^x_{k-1} \geq 0$. This is shown in Figure 4.3b for one realization of the reformulated UKF, applied to the non-augmented UKF algorithm proposed by Simon (2006), see also Kolås et al. (2008a). The performance in this case is excellent.

Divided Difference KF

The Divided Difference Kalman filter (DDKF) is based on polynomial approximations of the nonlinear transformations obtained with a multidimensional extension of Stirling's interpolation formula. DD1 is based on a first order approximation and DD2 is based on a second order approximation Nørgaard et al. (2000). As the UKF algorithms, the DDKF algorithms are Jacobian free algorithms and the mean and covariance are in principal calculated in the same manner as in the UKF - by some weighted sums. In van der Merwe (2004) it is shown that the DD2 KF has a slightly more accurate covariance estimate than the UKF based on the scaled unscented transform (SUT), but for all practical purposes there are no difference in estimation performance between the SUT based UKF and the DD2 KF van der Merwe (2004).

DD1 and DD2 KF are tested on the '2 state CSTR' case. The noise is assumed Gaussian. Note that the noise sequences used in the simulations for the DD1 KF and DD2 KF are not identical, and are also different than the noise sequence used in the previous section discussing EKF and UKF. The initial values and system parameters are the same as for the EKF and UKF simulations previously discussed. By constraining the corrected state estimate such that $\hat{x}_k \geq 0$, we get the results as given in Figure 4.4.

As Figure 4.4 shows, DD1 KF outperforms the EKF and converges to the true states, but not as fast as the constrained UKF algorithms discussed above. Also DD2 KF outperforms EKF and converge to the true states. The convergence speed is similar to the constrained non-augmented reformulated UKF algorithm discussed above.

4.6 Noise modeling concepts

As claimed earlier, the quality of the estimates may not only be dependent on the accuracy of the model and of the estimating method selected, but also of how process knowledge is applied. In our view, noise modeling has attracted more attention in the system identification literature (see

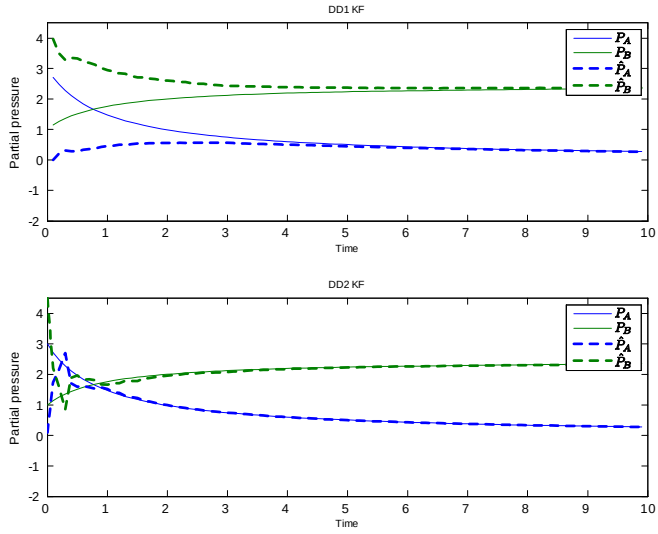

Figure 4.4: The simulation results using the DD1 KF (upper most subplot) and the DD2 KF (lower most subplot).

e.g. Ljung (1987)) than in the literature on identification and estimation for physics based models. One example is how white noise may enter an input-output model in different ways. Studying noise modeling based on a hypothesis that it is important to model noise correctly, implies in practice a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty. The 'additive noise paradigm' dominates textbooks and papers on recursive state estimation, i.e. Kalman filter type algorithms like the EKF and UKF (see e.g. Gelb (1974), Simon (2006), Søderstrøm and Stoica (1989)) even though uncertainty may enter a system in many different ways. The additive noise model structure is obviously reasonable in many applications. In others, however, this is not the case. One example are processes where control input uncertainty dominates, and where this noise depends on the value of the control input itself. It may for instance increase proportionally with the control input. A fruitful way to view noise modeling is to view this as a direct extension of the process of developing a model. We assume dynamic models which are developed using physical insight and process data, i.e. physics-based models. Having established and possibly validated such a model, it is at least in principle possible to quantify uncertainty. This may include uncertainty in initial conditions,

and in certain time-varying states and parameters, control inputs and measurements. Further, it may be possible to describe how noise enters the system, i.e. to structurally model how uncertainty affects the model.

Recursive state estimation algorithms usually assume that uncertainty enters through additive white noise sources. Further, unknown and time-varying parameters are often modeled similarly by augmenting the states with a parameter vector. Finally, initial model uncertainty is reflected through the choice of the initial covariance matrices for the states and parameters. In the sections to follow we investigate the effect of different system noise modeling methods.

4.6.1 Method 1 - Additive noise

Consider the gas-phase, reversible reaction as given by (4.4)-(4.7). Further we assume the reaction parameter k_r varies and is modelled as colored noise, and is estimated. The state and output vectors for the estimator are defined as

$$\hat{x} = \begin{bmatrix} \hat{P}_A \\ \hat{P}_B \\ \hat{k}_r \end{bmatrix} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix}, \qquad \hat{y} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \hat{x} \tag{4.8}$$

and extended with control inputs (u). Described the traditional way by assuming additive noise on the system we get:

$$\dot{\hat{x}}_1 = -2\hat{r} + u_1 + v_1 \tag{4.9}$$

$$\dot{\hat{x}}_2 = \hat{r} + u_2 + v_2 \tag{4.10}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{4.11}$$

where the state vector $\hat{x}$ is given by (4.8) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r \hat{x}_1^2 = \hat{x}_3 \hat{x}_1^2 \tag{4.12}$$

The noise vector is given by

$$v = \begin{bmatrix} v_1 & v_2 & v_{k_r} \end{bmatrix}^T \tag{4.13}$$

where $v_1 \sim N(0, \bar{v}_1^2)$, $v_2 \sim N(0, \bar{v}_2^2)$ and $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$ and the parameters by

$$\hat{\theta} = \hat{k}_r \tag{4.14}$$

Assume that the system (4.4)-(4.7) is considered as a semi-batch system in that the species B is removed and that the species A is refilled when a certain level of A is reached resulting in the control input described by (4.15)

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \left\{ \begin{array}{c} 4 - x_1) \\ -x_2 \end{array} \right\} & x_1 \leq 0.2 \\ \\ \left\{ \begin{array}{c} 0 \\ 0 \end{array} \right\} & otherwise \end{bmatrix} \tag{4.15}$$

The Jacobian H is given by the identity matrix I, and hence HQH^T in (4.1) is given by

$$HQH^T = Q = diag \begin{bmatrix} \bar{v}_1^2 & \bar{v}_2^2 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{4.16}$$

It is common practice to look at the state covariance as if it reflects the confidence in the system states $\hat{x}$, and that quantitatively a relatively high value means quite uncertain estimates and vice versa ((Søderstrøm and Stoica, 1989, p. 326)). A challenge using the proposed noise formulation might be that if the covariance has settled to a relative low value (i.e. certain state estimates) when the control inputs is applied, i.e. removing B and refilling A, this may introduce large errors in the estimates. Also worth noticing is that with this formulation there is no information about the control inputs in the covariance equation given by (4.1).

4.6.2 Method 2 - Noise in control inputs

We now consider that there is some uncertainty related to the control inputs, and that the uncertainty in the inputs can be expressed as a relative uncertainty. That is:

$$\dot{\hat{x}}_1 = -2\hat{r} + u_1(1 + v_{u_1}) \tag{4.17}$$

$$\dot{\hat{x}}_2 = \hat{r} + u_2(1 + v_{u_2}) \tag{4.18}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{4.19}$$

where the state vector $\hat{x}$ is given by (4.8) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r \hat{x}_1^2 = \hat{x}_3 \hat{x}_1^2 \tag{4.20}$$

The noise vector is given by

$$v = \begin{bmatrix} v_{u_1} & v_{u_2} & v_{k_r} \end{bmatrix}^T \tag{4.21}$$

where $v_{u_1} \sim N(0, \bar{v}_{u_1}^2)$, $v_2 \sim N(0, \bar{v}_{u_2}^2)$ and $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$ and the parameters by (4.14). The Jacobian H becomes

$$H = diag \begin{bmatrix} u_1 & u_2 & 1 \end{bmatrix} \tag{4.22}$$

Assuming the Q is described by

$$Q = diag \begin{bmatrix} \bar{v}_{u_1}^2 & \bar{v}_{u_2}^2 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{4.23}$$

the term HQH^T in (4.1) becomes

$$HQH^T = diag \begin{bmatrix} u_1^2\bar{v}_{u_1}^2 & u_2^2\bar{v}_{u_2}^2 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{4.24}$$

Consider again that the covariance reflects the uncertainty of the system states $\hat{x}$. In the case when the control inputs are applied, large errors in the estimates may be introduced. This is reflected in the proposed formulation (4.24) by the injection of the input uncertainty into the covariance function.

4.6.3 Method 3 - Noise in auxiliary variables

In this case the noise enters the system through the reaction rate r, instead of directly on the system states. That is:

$$\dot{\hat{x}}_1 = -2\hat{r} \tag{4.25}$$

$$\dot{\hat{x}}_2 = \hat{r} \tag{4.26}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{4.27}$$

where the state vector $\hat{x}$ is given by (4.8) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r\hat{x}_1^2 + v_r = \hat{x}_3\hat{x}_1^2 + v_r \tag{4.28}$$

The noise vector is given by

$$v = \begin{bmatrix} v_r & v_{k_r} \end{bmatrix}^T \tag{4.29}$$

where $v_r \sim N(0, \bar{v}_r^2)$ and $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$ and the parameters by (4.14). The system becomes

$$\dot{\hat{x}}_1 = -2\hat{k}_r\hat{x}_1^2 - 2v_r \tag{4.30}$$

$$\dot{\hat{x}}_2 = \hat{k}_r \hat{x}_1^2 + v_r \tag{4.31}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{4.32}$$

and the Jacobian H

$$H = \begin{bmatrix} -2 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.33}$$

Assuming the Q is described by

$$Q = diag \begin{bmatrix} \bar{v}_r^2 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{4.34}$$

the term HQH^T in (4.1) becomes

$$HQH^T = \begin{bmatrix} 4\bar{v}_r^2 & -2\bar{v}_r^2 & 0 \\ -2\bar{v}_r^2 & \bar{v}_r^2 & 0 \\ 0 & 0 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{4.35}$$

That is, by applying noise on the auxiliary variable $\hat{r}$, correlation is naturally introduced in the covariance calculation. This could also be seen as an alternative to off-diagonal tuning of Q in (4.16), in that the correlation enters the system naturally and correctly scaled in the off-diagonal elements.

4.6.4 Simulation results

2 state CSTR

We investigate the case as described by (4.4)-(4.8) using the UKF with constraint handling as in Chapter 5.2. Regarding the discrete system representation, the analytical model as given by Hasseltine and Rawlings (2003) is used. Further, the parameters for this system, if not otherwise noted, are $\Delta t = t_{k+1} - t_k = 0.1$, $P_0 = diag \begin{bmatrix} 6^2 & 6^2 & 0.015^2 \end{bmatrix}$, $k_r = 0.16$, $x_0 = \begin{bmatrix} 3 & 1 \end{bmatrix}^T$, $\hat{x}_0 = \begin{bmatrix} 0.1 & 4.5 & 0.9k_r \end{bmatrix}^T$. Note that the initial guess ($\hat{x}_0$) for the estimator is poor. The reaction parameter k_r is constant, but has a wrong initial estimate. The following constraints are applied to the UKF sigma points

$$\begin{aligned} Lower\ bounds: &\ [0, 0, 0.1]^T \\ Upper\ bounds: &\ [\infty,\ \infty,\ , 0.18]^T \end{aligned} \tag{4.36}$$

The state estimator used in this work is the fully augmented UKF with reformulation of the correction steps and the use of constraints as presented in Kolås et al. (2008a).

Method 1 The noise is modeled as in Method 1, the estimator constraints as (4.36) and the estimator tuning as

$$Q_k = diag\left[\begin{array}{cccc} \bar{v}_{k,1}^2 & \bar{v}_{k,2}^2 & \bar{v}_{k,k_r}^2 \end{array}\right] = diag\left[\begin{array}{ccc} \left(10^{-9}\right)^2 & \left(10^{-9}\right)^2 & \left(10^{-4}\right)^2 \end{array}\right]$$
$$R_k = \bar{w}_k^2 = 0.002^2 \tag{4.37}$$

In the case when the true process experiences no noise, we get the results as shown in Figure 4.5.

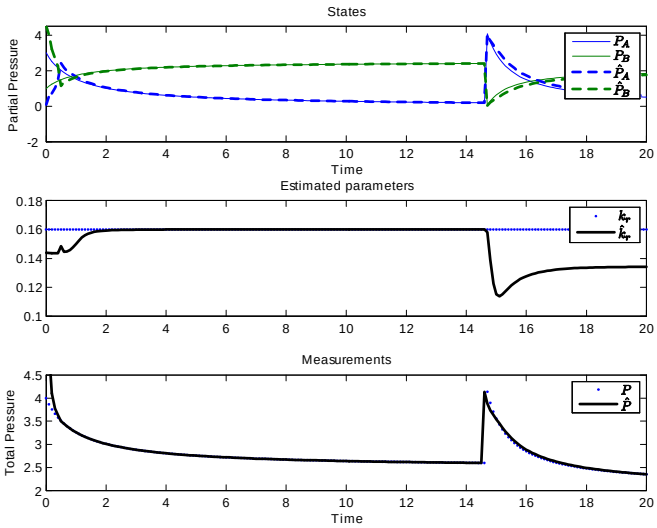

Figure 4.5: The figure shows the true and estimated states, the true and estimated reaction 'constant' k_r and the true and estimated output using Method 1. As the figure shows, the estimates of k_r is disturbed in the transient periode caused by the control input exitation.

As Figure 4.5 shows, the bad initial guess is handled very well, and the state estimate and output estimate is generally acceptable. However the parameter estimate during the transient period at approximately $t = 15$ is generally bad due to the control input excitation. Inspection of the time trace of the covariance is as expected, i.e. the control input excitation is not reflected in the covariance.

By modeling the noise by combining Method 1 and 2, and by that introduce uncertainty also in the control inputs, the system equations becomes

$$\dot{\hat{x}}_1 = -2\hat{r} + u_1(1 + v_{u_1}) + v_1 \tag{4.38}$$

$$\dot{\hat{x}}_2 = \hat{r} + u_2(1 + v_{u_2}) + v_2 \tag{4.39}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{4.40}$$

where the state vector $\hat{x}$ is given by (4.8) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r \hat{x}_1^2 = \hat{x}_3 \hat{x}_1^2 \tag{4.41}$$

The noise vector is given by $v = \begin{bmatrix} v_1 & v_2 & v_{k_r} & v_{u_1} & v_{u_2} \end{bmatrix}^T$ with $v_1 \sim N(0, \bar{v}_1^2)$, $v_2 \sim N(0, \bar{v}_2^2)$, $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$, $v_{u_1} \sim N(0, \bar{v}_{u_1}^2)$, $v_{u_2} \sim N(0, \bar{v}_{u_2}^2)$. HQH^T in (4.1) becomes

$$HQH^T = diag \begin{bmatrix} \bar{v}_1^2 + u_1^2 \bar{v}_{u_1}^2 & v_2^2 + u_2^2 \bar{v}_{u_2}^2 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{4.42}$$

With the estimator constraints as by (4.36), and the estimator tuning as

$$\begin{gathered} Q_k = diag \begin{bmatrix} \bar{v}_{k,1}^2 & \bar{v}_{k,2}^2 & \bar{v}_{k,k_r}^2 & \bar{v}_{k,u_1}^2 & \bar{v}_{k,u_2}^2 \end{bmatrix} = \\ diag \begin{bmatrix} \left(10^{-9}\right)^2 & \left(10^{-9}\right)^2 & \left(10^{-4}\right)^2 & 1^2 & 1^2 \end{bmatrix} \end{gathered} \tag{4.43}$$

$$R_k = \bar{w}_k^2 = 0.002^2$$

the results for the case when the simulator experience no noise, is as shown in Figure 4.6.

As Figure 4.6 shows, the estimate of the parameter k_r does not suffer from the control input excitation. However the 'cost' is some loss off accuracy after the control input excitation. Inspection of the time trace of the covariance shows that the control input excitation at approximately $t = 15$ is reflected in the covariance elements ($P_{x_k}(1,1)$, $P_{x_k}(1,2)$, $P_{x_k}(2,1)$, $P_{x_k}(2,2)$). That is, by applying the idea that the covariance should reflect the uncertainty in the states around the control input excitation, good results are achieved by combining the concept in Method 1 and 2 (see Kolås et al. (2008b) for further references).

4.7 Discussion

This paper centres on performance of different nonlinear state estimators and on noise modeling. Three different state estimators have been tested

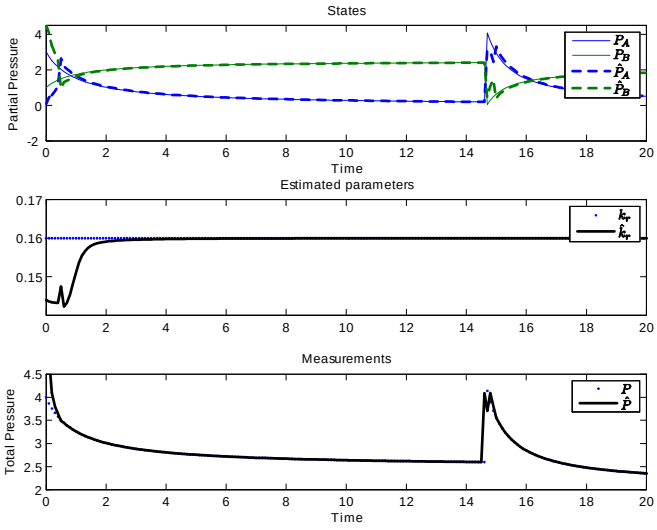

Figure 4.6: The figure shows the true and estimated states, the true and estimated reaction 'constant' k_r and the true and estimated output. As the figure shows the estimated k_r is not disturbed in the transient periode caused by the control input exitation.

on a simple and challenging problem showing quite different performance. This is not surprising. It does however indicate that the use of higher order filters may be required for demanding NMPC applications. Further, the use of constraint handling can be critical to obtain good estimator performance. This study is significantly extended in Kolås et al. (2008a). The results in this reference align with conclusions herein.

Noise modeling is significant for estimator performance. Again a simple example is used to illustrate this fact. A recent study Kolås et al. (2008b) expands on the results provided herein. In particular the conclusions coincide with the observations in this paper.

The results in this paper are consistent with our experience for NMPC applications in which the choice of nonlinear filter algorithms and the choice of reasonable noise model are critical factors to obtain good NMPC performance.

4.8 Conclusions

Our experience shows that the state estimation problem is indeed the critical challenge in most NMPC applications. Furthermore, practice reveals that the use of higher order filters like the UKF is important to obtain good performance in challenging applications. Finally, alternative noise models to the dominating 'additive noise paradigm' has a positive impact on estimation performance in some cases.

4.9 Acknowledgements

The financial support of Hydro Aluminium AS is gratefully acknowledged.

Chapter 5

Constrained nonlinear state estimation based on the UKF approach

This chapter is based on (Kolås, Foss and Schei 2008) as submitted[1] *to Computers and Chemical Engineering 2008.*

5.1 Abstract

In this paper we investigate the use of an alternative to the Extended Kalman filter (EKF), the Unscented Kalman filter (UKF). First we will give a broad overview of different UKF algorithms, then present an extension to the ensemble of UKF algorithms, and finally address the issue of how to add constraints using the UKF approach. The performance of the constrained approach is compared with EKF and a selection of UKF algorithms on nonlinear process systems with multimodal probability density functions. The conclusion is that with an algebraic reformulation of the correction part, the reformulated UKF shows very strong performance on our selection of nonlinear constrained process systems.

Keywords: Unscented Kalman filter, Extended Kalman filter, Nonlinear state estimation, Constrained nonlinear systems

[1] Resubmitted version based on feedback from reviewers.

5.2 Introduction

Nonlinear state estimation is a broad field[2] which includes many algorithms such as Moving Horizon Estimation, the Particle filter, the Ensemble Kalman filter, the Unscented Kalman filter and the Extended Kalman filter, just to mention some. The Extended Kalman filter (EKF), which was originally proposed by Stanley Schmidth in 1967 (Bellatoni and Dodge (1967)) in order to apply the Kalman filter to nonlinear spacecraft navigation problems, is probably the most widely used method in applied nonlinear state estimation. However, several authors have experienced shortcomings applying the EKF to systems with severe nonlinearity and/or constraints (see e.g. Julier and Uhlmann (1994), Julier et al. (1995), Schei (1997), Nørgaard et al. (2000), Rao (2000), Hasseltine and Rawlings (2003), Bizup and Brown (2003), Chen et al. (2006), Vachhani et al. (2006) and Kandepu et al. (2008)). The shortcomings are related to difficulties in determining the Jacobians, errors introduced by linearization and/or the ability to deal with systems with multimodal or asymmetric probability density functions (pdf).

State estimation may introduce some challenges in systems based on first principles. In e.g. chemical processes one may wish to estimate concentrations. Mathematically the model may allow negative concentration, whilst this is not physically possible. A first principle model is also often nonlinear, and theory based on Gaussian noise may not be applicable, since Gaussian noise propagated through a nonlinear model is no longer Gaussian. Further, nonlinearities may be severe such that theory based on linearizations are too inaccurate and provide poor estimation accuracy. At last the nonlinear system may have a skew or multimodal probability density function (pdf). If $\hat{x}_k$ is an estimate of concentrations in a physical system, we may introduce constraints in order to force the estimates to have physical meaning. Hence the use of constraints may be important.

If handling of constraints is unavoidable, the EKF has some limitations in propagating the constraints both through the state and covariance calculations. To overcome limitations in the EKF, several approaches and alternatives have been suggested. In this paper, we focus on the Unscented Kalman filter (UKF). In particular several variants of the UKF algorithm with constraint handling are investigated on cases where the EKF has shown bad performance. It is, however, outside the scope of the paper to include an in-depth discussion on MHE and Particle filters beyond the fact that we

[2] For a review of nonlinear state estimation see Daum (1986) and Daum (2005).

use examples on which such algorithms have been evaluated, see Hasseltine and Rawlings (2003) and Rawlings and Bakshi (2006).

Literature discussing constraint implementation in the UKF approach is rather limited. In Julier and Uhlmann (1994) the idea of including constraints in the UKF approach is in fact listed as a general possibility, but without being further discussed. The main contributions on constraint handling are Vachhani et al. (2006) and Kandepu et al. (2008). Further, Li and Leung (2004) use equality constraints on the corrected estimate, and Julier and Laviola (2007) discuss two methods for nonlinear equality constraint handling.

This paper is organized as follows. First we introduce the Kalman filter-philosophy based on the work by Rudolph Kalman (Kalman (1960)), before we introduce the reader to the UKF and present several algorithms based on this approach. Subsequently we propose an extension to the UKF framework by reformulating the correction step. This reformulation, combined with different constraint handling techniques, results in good performance on the selected cases. We also show that the selection of the square root algorithm may be of importance when it comes to convergence speed of the UKF. Further, we extend the work of Vachhani et al. (2006) to Nonlinear Programming (NLP) UKF by presenting the full state estimation algorithm for alternative formulations of the UKF. Moreover, we show how this algorithm can be simplified to a Quadratic Program (QP) UKF given some additional system constraints. Various simulation examples (taken from Hasseltine and Rawlings (2003)), are used to show how a selection of the algorithms performs before a discussion and conclusions end the paper.

5.3 System description

In this work we address the general discrete time nonlinear system given by the state space formulation

$$\begin{aligned} x_k &= f(x_{k-1}, u_{k-1}, v_{k-1}) \\ & \quad x_0 \text{ - given} \end{aligned} \tag{5.1}$$

$$y_{k-1} = h(x_{k-1}, w_{k-1}) \tag{5.2}$$

where x_k denotes the system states, y_k the model output (or measurements), u_k deterministic control inputs, v_k system state noise input (sta-

tionary or time-varying stochastic), w_k output (measurement) noise (stationary or time-varying stochastic), and $f(\cdot)$ and $h(\cdot)$ are nonlinear functions. Note that $f(\cdot)$ can also be considered as the integration of the continous-time transition function over a unit sample-time interval (see e.g. Vachhani et al. (2006)). All variables are vectors with appropriate dimensions. Linear systems are given by

$$f(x_k, u_k, v_k) = A_k x_k + B_k u_k + C_k v_k \tag{5.3}$$

$$h(x_k, w_k) = D_k x_k + w_k \tag{5.4}$$

covering both linear time-varying (LTV) systems and linear time-invariant (LTI) systems. A_k, B_k, C_k and D_k are matrices with appropriate dimensions describing the linear relationship between the variables. Estimated values are denoted with a $\hat{}$, typically $\hat{x}_k$, $\hat{y}_k$ etc. Further, the simulator and the estimator share the same model, and hence there is no modelling error, although model errors are introduced through the noise inputs v_k and w_k.

5.4 The Kalman-approach to state estimation

In 1960 Rudolph Kalman (Kalman (1960)) presented a recursive state estimation method, which has become known as the Kalman filter. The important contribution was the fact that a recursive algorithm could be used to accurately compute the first and second order moments (mean and covariance) of a linear system corrupted by Gaussian white noise on its system and output models. However, it may be worth mentioning that Kalman did not restrict the Kalman filter to linear time invariant systems with Gaussian distributions, nor did he assume any specific probability distribution (Julier and Uhlmann (2004)). Kalman's assumptions were that the system random variables could be consistently estimated by sequentially updating the first and second order moments (mean and covariance), and that the estimator was on the linear form

$$\hat{x}_k = \hat{x}_k^- + K_k \cdot e_k \tag{5.5}$$

The Kalman filter consists of two parts; a forward prediction part and a correction part. The prediction part computes a predicted (a prior) estimate of the first and second order moments at time k given information

up until $k-1$ denoted by $\hat{x}_k^-$ and $P_{x_k}^-$. The correction part computes the corrected (posterior) estimates $\hat{x}_k$ and P_{x_k} using available data at time k. In (5.5) the predicted estimate ($\hat{x}_k^-$) is updated by a linear gain (K_k) times an error (e_k). The error part is the deviation between an output value(s) (y_k) and the prediction of the said output(s) ($\hat{y}_k$). Hence, (5.5) can be formulated as

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - \hat{y}_k) \tag{5.6}$$

where the linear gain (K_k), or the so-called Kalman gain, which for the

algorithms discussed in this paper, can be shown to yield the solution

$$K_k = P_{x_k y_k} P_{y_k y_k}^{-1} \tag{5.7}$$

where $P_{x_k y_k}$ and $P_{y_k y_k}$ are the cross covariance and output covariance, respectively .

All Kalman filters do follow the prediction-correction structure. As a vehicle to compare alternative algorithms, we have divided the algorithm into steps as shown in Table 5.1. In the tables, used throughout this paper to describe the algorithms, we have chosen to keep the column labeled 'Algorithm step' the same for all algorithms, although all steps do not appear in all algorithms. This is because we think this will help in comparing the steps involved in each algorithm. If a step is not applicable to an algorithm, it is indicated by an '-' in the right most column.

Note that all algorithms must be initialized with first and second order moment state information. The prediction and correction part of an algorithm is covered by steps 1-6 and steps 7-12, respectively.

5.5 Recursive state estimators

The Kalman-filter structure is widely applied for nonlinear state estimation. The subsequent sections describe the state estimators studied in this work. First the Kalman filter (KF) and the Extended Kalman filter (EKF) are introduced to serve as a comparison for the Jacobian free algorithms to follow. Note that the formulation of both the KF and the EKF may be unfamiliar, since several algebraic steps are kept in order to show the 1:1 relationship of the algorithms. This results in seemingly more computational steps than in the common formulation of KF/EKF.

Table 5.1: The table explains the meaning of each step used throught the paper

Step no.	Algorithm step	Meaning
1	χ_{k-1}	Sigma points (see Chapter 4.3)
2	$\chi^{x-}_{k,i}$	Propagated state sigma points
3	$\hat{x}^-_k$	Predicted state estimate
4	$P^-_{x_k}$	Predicted covariance
5	$\gamma_{k,i}$	Propagated output sigma points
6	$\widehat{y}_k$	Predicted output
7	$P_{y_k y_k}$	Output Covariance
8	$P_{x_k y_k}$	Cross covariance
9	K_k	Kalman gain
10	$\chi^x_{k,i}$	Corrected sigma points
11	$\hat{x}_k$	Corrected state estimate
12	P_{x_k}	Corrected covariance

5.5.1 Kalman filter

If one assume the linear time invariant (LTV) system given by (5.3) and (5.4) with the system covariance Q_k and output covariance R_k and corrupted by Gaussian noise $v_k = N(\bar{v}, Q)$, $w_k = N(\bar{w}, R)$, $(\bar{v} = \bar{w} = 0)$, the Kalman filter approximates the said system as given by Table 5.2.

Table 5.2: The KF algorithm

χ_{k-1}	-
$\chi^{x-}_{k,i}$	-
$\widehat{x}^-_k =$	$A_{k-1}\hat{x}_{k-1} + B_{k-1}u_{k-1} + C_{k-1}v_{k-1}$
$P^-_{x_k} =$	$A_{k-1}P_{\hat{x}_{k-1}}A^T_{k-1} + D_{k-1}Q_{k-1}D^T_{k-1}$
$\gamma_{k,i}$	-
$\widehat{y}_k =$	$D_{k-1}\hat{x}_{k-1} + w_{k-1}$
$P_{y_k y_k} =$	$D_k P^-_{x_k} D^T_k + R_k$
$P_{x_k y_k} =$	$P^-_{x_k} D^T_k$
$K_k =$	$P_{x_k y_k} P^{-1}_{y_k y_k}$
$\chi^x_{k,i}$	-
$\widehat{x}_k =$	$\widehat{x}^-_k + K_k(y_k - \widehat{y}_k)$
$P_{x_k} =$	$(I - K_k D)P^-_{x_k}$

The EKF is based on (cf Table 5.3) the nonlinear system given by (5.1)

- (5.2) with the same noise models as for the KF.

Table 5.3: The EKF algorithm

χ_{k-1}	-
$\chi_{k,i}^{x-}$	Calc Jacobians $\nabla f_{\hat{x}_{k-1}}$,$\nabla f_{v_{k-1}}$
$\hat{x}_k^- =$	$f(\hat{x}_{k-1}, u_{k-1}, v_{k-1})$
$P_{x_k}^- =$	$\nabla f_{\hat{x}_{k-1}} P_{\hat{x}_{k-1}}^- \nabla f_{\hat{x}_{k-1}}^T + \nabla f_{v_{k-1}} Q_{k-1} \nabla f_{v_{k-1}}^T$
$\gamma_{k,i}$	Calc Jacobians $\nabla h_{\hat{x}_k^-}$,∇h_{w_k}
$\hat{y}_k =$	$h(\hat{x}_k^-, w_k)$
$P_{y_k y_k} =$	$\nabla h_{\hat{x}_k^-} P_{x_k}^- \nabla h_{\hat{x}_k^-}^T + \nabla h_{w_k} R_k \nabla h_{w_k}^T$
$P_{x_k y_k} =$	$P_{x_k}^- \nabla h_{x_k}^T$
$K_k =$	$P_{x_k y_k} P_{y_k y_k}^{-1}$
$\chi_{k,i}^{x}$	-
$\hat{x}_k =$	$\hat{x}_k^- + K_k(y_k - \hat{y}_k)$
$P_{x_k} =$	$(I - K_k \nabla h_{\hat{x}_k^-}) P_{x_k}^-$

One should notice the slightly different definition of the predicted output between the KF and the EKF. We have used the Jacobians

$$\begin{aligned}
\nabla f_{\hat{x}_{k-1}} &= \frac{\partial}{\partial x} f(x,u,v)|_{x=\hat{x}_{k-1},\, u=u_{k-1},\, v=v_{k-1}} \\
\nabla f_{v_{k-1}} &= \frac{\partial}{\partial v} f(x,u,v)|_{x=\hat{x}_{k-1},\, u=u_{k-1},\, v=v_{k-1}} \\
\nabla h_{\hat{x}_k^-} &= \frac{\partial}{\partial x} h(x,w)|_{x=\hat{x}_k^-,\, w=w_{k-1}} \\
\nabla h_{w_k} &= \frac{\partial}{\partial w} h(x,w)|_{x=\hat{x}_k^-,\, w=w_{k-1}}
\end{aligned}$$

Since the Jacobian $\nabla h_{\hat{x}_k^-}$and $\nabla h_{w_k}$are calculated around $\hat{x}_k^-$ (the predicted state estimate) this yields a difference since the KF uses $\hat{x}_{k-1}$(the corrected state estimate). Note also that we have placed the calculation of the Jacobians in the steps $\chi_{k,i}^{x-}$ and $\gamma_{k,i}$ to illustrate that the calculation of the Jacobians, with respect to performance, can be compared to the sigma points step used in the UKF algorithms presented later.

5.5.2 The Unscented Kalman filter (UKF)

The Unscented Kalman[3] filter, proposed by Julier and Uhlmann (1994)[4], is based on the intuition that "*it is easier to approximate a probability distribution than a nonlinear function*" (Julier et al. (2000), Julier (2002)). By using the Unscented transform to compute the mean and covariance, the Unscented Kalman filter avoids the need to use Jacobians in the algorithm. The probability distribution is approximated by a set of deterministic points which captures the mean and covariance of the distribution. These points, called sigma points, are then processed through the nonlinear model of the system, producing a set of propagated sigma points. By choosing appropriate weights, the weighted average and the weighted outer product of the transformed points provides an estimate of the mean (for example $\hat{x}_k^-$) and covariance (for example $P_{x_k}^-$) of the transformed distribution. To elaborate, a sigma point set, χ_k^x is a matrix which contains states or sigma points $\chi_k^x = \left[\begin{array}{ccc} \chi_{k,1}^{x^T} & \cdots & \chi_{k,l}^{x^T} \end{array} \right]$, where l typically equals $2n+1$ if $x_k \in R^n$.

Passing the vector of sigma points $\chi_{k-1,i}^x$ through a nonlinear function results in the propagated sigma points

$$\chi_{k,i}^x = g(\chi_{k-1,i}^x) \tag{5.8}$$

and the approximated mean $\hat{x}_k$ and the covariance P_{x_k} can be calculated as

$$\hat{x}_k = \sum_{i=1}^{2n} W_i^x \cdot \chi_{k,i}^x \tag{5.9}$$

$$P_{x_k} = \sum_{i=1}^{2n} W_i^c (\chi_{k,i}^x - \hat{x}_k)(\chi_{k,i}^x - \hat{x}_k)^T{}_k \tag{5.10}$$

Several UKF algorithms have been proposed (for an overview see van der Merwe (2004)) and they differ mainly in how the noise (system noise and output noise) is injected into the system, how the set of sigma point is selected and how the weights are calculated. There are also UKF algorithms based on square root formulations (see van der Merwe (2004)) and UKF

[3] In the original work of Julier and Uhlmann (1994) the Unscented Kalman filter was named 'The New Filter'.

[4] We refer to this document as the first work of UKF, since it was submitted to IEEE Transactions on Automatic Control in 1994, although not accepted before in 2000.

algorithms addressing nonlinear state model and linear output (measurement) model (also known as Rao-Blackwellized UKF's) (Briers et al. (2003), Hao et al. (2007)). Neither the square root UKF nor the Rao-Blackwellized UKF algorithms have our attention in this work.

In the next chapters we will present several UKF algorithms, but before they are presented, some words about the selection of the sigma points and the weights may be appropriate.

In the original work by Julier and Uhlmann (1994) a symmetric sigma point set was proposed, resulting in $2n + 1$ sigma points where $x_k \in R^n$. However, the calculation of the sigma points is considered the most computational demanding steps in the UKF, resulting in an effort to reduce the number of sigma points. This resulted in the 'Simplex Sigma point set'/'Minimal skew simplex sigma point set' (Julier and Uhlmann (2002), Julier and Uhlmann (2004)) , the 'Spherical simplex sigma point set' (Julier (2003)) and the 'Higher order unscented filter' (Tenne and Singh (2003)). Throughout this paper we are using the symmetric sigma point scheme as a basis.

Having the mean and covariance available, the symmetric sigma point set χ_k is selected by constructing the matrix

$$\chi_k = \left[\begin{array}{cc} \hat{x}_k + \Gamma\sqrt{P_{x_k}} & \hat{x}_k - \Gamma\sqrt{P_{x_k}} \end{array} \right] \tag{5.11}$$

where Γ is a scaling factor. Note also that the notation $\hat{x}_k \pm \Gamma\sqrt{P_{x_k}}$ means that the vector $\hat{x}_k$ is added/subtracted to each column in the matrix $\Gamma\sqrt{P_{x_k}}$. In Julier and Uhlmann (1994) they added the freedom to include in χ_k any multiples of the mean $\hat{x}_k$, since it would not affect the mean, only the scaling factor Γ (Julier and Uhlmann (1994)) leading to the sigma point calculation

$$\chi_k = \left[\begin{array}{ccc} \hat{x}_k & \hat{x}_k + \Gamma\sqrt{P_{x_k}} & \hat{x}_k - \Gamma\sqrt{P_{x_k}} \end{array} \right] \tag{5.12}$$

In the UKF literature, (5.12) is the most applied method in selecting the sigma point set. We will later introduce an UKF algorithm based on the sigma point selection given by (5.11). Further, in the mentioned literature, three different ways of calculating the weights are proposed. We will, however, try to merge these into one set of equations and suggest different sets of parameters in order to achieve each of them. From this, a general scaling factor Γ can be formulated as

$$\Gamma = \sqrt{n + \lambda} \tag{5.13}$$

where

$$\lambda = \kappa \tag{5.14}$$

is suggested in the original work of Julier and Uhlmann (1994). Julier and Uhlmann (1994) and Julier et al. (2000) proposed

$$\kappa = 3 - n \tag{5.15}$$

if the system is Gaussian[5]. However care should be taken selecting κ, since $\kappa < 0$ may lead to negative definite covariance calculations.

Motivated by making the scales independent of the system size, the Scaled Unscented Transform was introduced (Julier (2002)), and further extended by van der Merwe (2004) suggesting

$$\lambda = \alpha^2 (n + \kappa) - n \tag{5.16}$$

The weights $W_i^x$and W_i^c are calculated as

$$\begin{aligned} W_0^x &= \lambda/(n+\lambda) \\ W_0^c &= \lambda/(n+\lambda) + (1 - \alpha^2 + \beta) \\ W_i^x &= W_i^c = 1/\left(2(n+\lambda)\right) \end{aligned} \tag{5.17}$$

van der Merwe (2004) propose

$$\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \tag{5.18}$$

if the system is Gaussian. Note that with

$$\begin{bmatrix} \alpha & \beta \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{5.19}$$

one achieve the original proposed scaling factor and weights as proposed in the original work of Julier and Uhlmann (1994). When using (5.11) as the basis for the sigma point selection

$$\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{5.20}$$

should be used according to (Simon (2006)) in applying (5.17).

[5]It can be shown that $(n + \kappa) = 3$ minimizes the difference between the moments in the Taylor series of the standard Gaussian and the sigma points up to the fourth order (Julier et al. (2000)).

Calculating the sigma points involves computing the matrix square root[6] of $\sqrt{P_{x_k}}$, and any matrix square root method can be used (Julier and Uhlmann (2004)). The Cholesky decomposition is the preferred method, producing a lower triangular matrix. However, one should also consider using the symmetric square root (producing a symmetric matrix), since the choice of matrix square root affects the errors in the higher order terms by adjusting the way in which the errors between the approximated Taylor series and the true Taylor series are distributed among the different states (Julier et al. (2000)). As will be demonstrated in the simulation chapter, this may affect the convergence performance.

UKF with additive noise

Assume the system given by (5.1) - (5.2) with the system covariance Q_k and output covariance R_k and corrupted by additive Gaussian noise $v_k = N(\bar{v}, Q)$, $w_k = N(\bar{w}, R)$. The UKF for this system is given by Table 5.4. Note that $\chi_{k-1} = \chi^x_{k-1}$. Note also that the notation $\hat{x}_{k-1} \pm \Gamma\sqrt{P_{x_{k-1}}}$ means that the vector $\hat{x}_{k-1}$ is added/subtracted to each column in the matrix $\Gamma\sqrt{P_{x_{k-1}}}$.

The algorithm in Table 5.4 uses (5.11) in selecting the sigma point set χ_{k-1} and an optional step using the updated propagated sigma points calculated as $\chi^{x-}_k = \left[\hat{x}^-_k + \Gamma\sqrt{P^-_{x_k}} \quad \hat{x}^-_k - \Gamma\sqrt{P^-_{x_k}}\right]$ before propagating $\chi^{x-}_{k,i}$ through the output function calculating $\gamma_{k,i}$.

In Julier et al. (1995), (5.12) is used when selecting the sigma point set, thus leading to $2n + 1$ sigma points, opposed to $2n$ sigma points used in the algorithm described in Table 5.4.

Based on the work by Julier et al. (1995) and van der Merwe (2004) the algorithm with augmented sigma points is as shown in Table 5.5.

As in the algorithm by in Table 5.4, an updated sigma point set can be calculated before propagating the sigma points through the output function. van der Merwe (2004) proposes an alternative method, and claim that using the updated sigma points removes any third order moments from the original propagated sigma point set $\chi^{x-}_{k,i}$. Hence, if this is a problem, one may consider the method proposed by van der Merwe (2004).

[6] Matrix square root is a method such that $P = \left(\sqrt{P}\right)^T \sqrt{P}$

Table 5.4: The UKF algorithm proposed by Simon (2006). Note the noise is assumed additive.

$$
\begin{aligned}
\chi_{k-1} &= \left[\begin{array}{cc} \hat{x}_{k-1} + \Gamma\sqrt{P_{x_{k-1}}} & \hat{x}_{k-1} - \Gamma\sqrt{P_{x_{k-1}}} \end{array} \right] \\
\chi^{x-}_{k,i} &= f(\chi^{x}_{k-1,i}, u_{k-1}) \\
\hat{x}^{-}_{k} &= \sum_{i=1}^{2n} W^{x}_{i} \cdot \chi^{x-}_{k,i} \\
P^{-}_{x_k} &= \sum_{i=1}^{2n} W^{c}_{i} (\chi^{x-}_{k,i} - \hat{x}^{-}_{k})(\chi^{x-}_{k,i} - \hat{x}^{-}_{k})^{T} + Q_k \\
\gamma_{k,i} &= h(\chi^{x-}_{k,i}) \\
\hat{y}_k &= \sum_{i=1}^{2n} W^{x}_{i} \cdot \gamma_{k,i} \\
P_{y_k y_k} &= \sum_{i=1}^{2n} W^{c}_{i} (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^{T} + R_k \\
P_{x_k y_k} &= \sum_{i=1}^{2n} W^{c}_{i} (\chi^{x-}_{k,i} - \hat{x}^{-}_{k})(\gamma_{k,i} - \hat{y}_k)^{T} \\
K_k &= P_{x_k y_k} P^{-1}_{y_k y_k} \\
\chi^{x}_{k,i} &= \text{-} \\
\hat{x}_k &= \hat{x}^{-}_{k} + K_k(y_k - \hat{y}_k) \\
P_{x_k} &= P^{-}_{x_k} - K_k P_{y_k y_k} K^{T}_{k}
\end{aligned}
$$

Table 5.5: The UKF algorithm with additive noise (Julier/van der Merwe)

$$
\begin{aligned}
\chi_{k-1} &= \begin{bmatrix} \hat{x}_{k-1} & \hat{x}_{k-1} + \Gamma\sqrt{P_{x_{k-1}}} & \hat{x}_{k-1} - \Gamma\sqrt{P_{x_{k-1}}} \end{bmatrix} \\
\chi^{x-}_{k,i} &= f(\chi^{x}_{k-1,i}, u_{k-1}) \\
\hat{x}^{-}_{k} &= \sum_{i=0}^{2n} W^{x}_{i} \cdot \chi^{x-}_{k,i} \\
P^{-}_{x_k} &= \sum_{i=0}^{2n} W^{c}_{i} (\chi^{x-}_{k,i} - \hat{x}^{-}_{k})(\chi^{x-}_{k,i} - \hat{x}^{-}_{k})^{T} + Q_k \\
\gamma_{k,i} &= h(\chi^{x-}_{k,i}) \\
\hat{y}_k &= \sum_{i=0}^{2n} W^{x}_{i} \cdot \gamma_{k,i} \\
P_{y_k y_k} &= \sum_{i=0}^{2n} W^{c}_{i} (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^{T} + R_k \\
P_{x_k y_k} &= \sum_{i=0}^{2n} W^{c}_{i} (\chi^{x-}_{k,i} - \hat{x}^{-}_{k})(\gamma_{k,i} - \hat{y}_k)^{T} \\
K_k &= P_{x_k y_k} P^{-1}_{y_k y_k} \\
\chi^{x}_{k,i} &\quad \text{-} \\
\hat{x}_k &= \hat{x}^{-}_{k} + K_k(y_k - \hat{y}_k) \\
P_{x_k} &= P^{-}_{x_k} - K_k P_{y_k y_k} K^{T}_{k}
\end{aligned}
$$

UKF with augmented system noise

The same system as in the previous chapter is considered. In the original work of Julier and Uhlmann (1994), the system noise is not assumed additive, but is augmented into the sigma point set. The reason is that by augmenting the assumed system noise into the sigma point set, the noise is also accounted for in the mean value (Julier and Uhlmann (1994)). The augmentation is done by merging the system noise covariance matrix Q_k and the covariance matrix P_{x_k} such that

$$P^a = \begin{bmatrix} P_{x_{k-1}} & 0 \\ 0 & Q_{k-1} \end{bmatrix} \tag{5.21}$$

This produces more sigma points, and Zandt (2001) demonstrates that increasing the number of sigma points may increase the accuracy of the UKF algorithm, however at the expense of increased computational cost. The UKF with augmented system noise that approximates the said system is given by Table 5.6. Note that the sigma points $\chi_{k-1} = [\chi^{T^x}_{k-1} \quad \chi^{T^v}_{k-1}]^T$ and the augmented state vector $\hat{x}^a_{k-1} = [\hat{x}^T_{k-1} \; 0]^T$. Note also that the notation $\hat{x}^a_{k-1} \pm \Gamma\sqrt{P^a}$ means that the vector $\hat{x}^a_{k-1}$ is added/subtracted to each column of $\Gamma\sqrt{P^a}$.

Fully augmented UKF

As in the previous sections, we consider the nonlinear system given by (5.1) - (5.2) corrupted by Gaussian noise $v_k = N(\bar{v}, Q)$, $w_k = N(\bar{w}, R)$. In the work of van der Merwe (2004), both the system noise and the output noise is augmented into the sigma point set. This is done by merging the system noise covariance matrix Q_k and the output noise covariance matrix R_k with the covariance matrix P_{x_k} such that

$$P^a = \begin{bmatrix} P_{x_{k-1}} & 0 & 0 \\ 0 & Q_{k-1} & 0 \\ 0 & 0 & R_{k-1} \end{bmatrix} \tag{5.22}$$

The UKF is shown in Table 5.7. Note that $\chi_{k-1} = [\chi^{T^x}_{k-1} \quad \chi^{T^v}_{k-1} \quad \chi^{T^w}_{k-1}]^T$ and the augmented state vector $\hat{x}^a_{k-1} = \begin{bmatrix} \hat{x}^T_{k-1} \; 0 \; 0 \end{bmatrix}^T$.

Table 5.6: The UKF algorithm with augmented process noise (Julier 1994)

$$
\begin{aligned}
\chi_{k-1} &= \begin{bmatrix} \hat{x}^a_{k-1} & \hat{x}^a_{k-1} + \Gamma\sqrt{P^a} & \hat{x}^a_{k-1} - \Gamma\sqrt{P^a} \end{bmatrix} \\
\chi^{x-}_{k,i} &= f(\chi^x_{k-1,i}, u_{k-1}, \chi^v_{k-1,i}) \\
\hat{x}^-_k &= \sum_{i=0}^{2n} W^x_i \cdot \chi^{x-}_{k,i} \\
P^-_{x_k} &= \sum_{i=0}^{2n} W^c_i (\chi^{x-}_{k,i} - \hat{x}^-_k)(\chi^{x-}_{k,i} - \hat{x}^-_k)^T \\
\gamma_{k,i} &= h(\chi^{x-}_{k,i}) \\
\hat{y}_k &= \sum_{i=0}^{2n} W^x_i \cdot \gamma_{k,i} \\
P_{y_k y_k} &= \sum_{i=0}^{2n} W^c_i (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^T + R_k \\
P_{x_k y_k} &= \sum_{i=0}^{2n} W^c_i (\chi^{x-}_{k,i} - \hat{x}^-_k)(\gamma_{k,i} - \hat{y}_k)^T \\
K_k &= P_{x_k y_k} P^{-1}_{y_k y_k} \\
\chi^x_{k,i} &= \text{-} \\
\hat{x}_k &= \hat{x}^-_k + K_k(y_k - \hat{y}_k) \\
P_{x_k} &= P^-_{x_k} - K_k P_{y_k y_k} K^T_k
\end{aligned}
$$

Table 5.7: The fully augmented UKF algorithm (van der Merwe 2000)

$$
\begin{aligned}
\chi_{k-1} &= \left[\begin{array}{ccc} \hat{x}^a_{k-1} & \hat{x}^a_{k-1} + \Gamma\sqrt{P^a} & \hat{x}^a_{k-1} - \Gamma\sqrt{P^a} \end{array}\right] \\
\chi^{x-}_{k,i} &= f(\chi^x_{k-1,i}, u_{k-1}, \chi^v_{k-1,i}) \\
\hat{x}^-_k &= \sum_{i=0}^{2n} W^x_i \cdot \chi^{x-}_{k,i} \\
P^-_{x_k} &= \sum_{i=0}^{2n} W^c_i (\chi^{x-}_{k,i} - \hat{x}^-_k)(\chi^{x-}_{k,i} - \hat{x}^-_k)^T \\
\gamma_{k,i} &= h(\chi^{x-}_{k,i}, \chi^w_{k,i}) \\
\hat{y}_k &= \sum_{i=0}^{2n} W^x_i \cdot \gamma_{k,i} \\
P_{y_k y_k} &= \sum_{i=0}^{2n} W^c_i (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^T \\
P_{x_k y_k} &= \sum_{i=0}^{2n} W^c_i (\chi^{x-}_{k,i} - \hat{x}^-_k)(\gamma_{k,i} - \hat{y}_k)^T \\
K_k &= P_{x_k y_k} P^{-1}_{y_k y_k} \\
\chi^x_{k,i} &= \text{-} \\
\hat{x}_k &= \hat{x}^-_k + K_k(y_k - \hat{y}_k) \\
P_{x_k} &= P^-_{x_k} - K_k P_{y_k y_k} K^T_k
\end{aligned}
$$

The reformulated UKF

Motivated by improved numerical behavior and flexible constraint handling, we propose an alternative UKF-formulation for the nonlinear system with noise inputs (assuming noise corrupted by Gaussian noise). By focusing on the sigma points and defining

$$\chi^x_{k,i} = \chi^{x^-}_{k,i} + K_k(y_k - h(\chi^{x^-}_{k,i}, \chi^w_{k,i})) \tag{5.23}$$

it can be shown that

$$\hat{x}_k = \sum_{i=0}^{2n} W^x_i \cdot \chi^x_{k,i} = \hat{x}^-_k + K_k(y_k - \hat{y}_k) \tag{5.24}$$

and that

$$P_{x_k} = \sum_{i=0}^{2n} W^c_i(\chi^x_{k,i} - \hat{x}_k)(\chi^x_{k,i} - \hat{x}_k)^T = P^-_{x_k} - K_k P_{y_k y_k} K^T_k \tag{5.25}$$

A proof is provided in Appendix A. The augmented (based on (5.22)) reformulated UKF is given by Table 5.8.

Table 5.8: The fully augmented UKF algorithm with reformulated correction steps

$$
\begin{aligned}
\chi_{k-1} &= \begin{bmatrix} \hat{x}^a_{k-1} & \hat{x}^a_{k-1} + \Gamma\sqrt{P^a} & \hat{x}^a_{k-1} - \Gamma\sqrt{P^a} \end{bmatrix} \\
\chi^{x-}_{k,i} &= f(\chi^x_{k-1,i}, u_{k-1}, \chi^v_{k-1,i}) \\
\hat{x}^-_k &= \sum_{i=0}^{2n} W^x_i \cdot \chi^{x-}_{k,i} \\
P^-_{x_k} &= \sum_{i=0}^{2n} W^c_i (\chi^{x-}_{k,i} - \hat{x}^-_k)(\chi^{x-}_{k,i} - \hat{x}^-_k)^T \\
\gamma_{k,i} &= h(\chi^{x-}_{k,i}, \chi^w_{k,i}) \\
\hat{y}_k &= \sum_{i=0}^{2n} W^x_i \cdot \gamma_{k,i} \\
P_{y_k y_k} &= \sum_{i=0}^{2n} W^c_i (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^T \\
P_{x_k y_k} &= \sum_{i=0}^{2n} W^c_i (\chi^{x-}_{k,i} - \hat{x}^-_k)(\gamma_{k,i} - \hat{y}_k)^T \\
K_k &= P_{x_k y_k} P^{-1}_{y_k y_k} \\
\chi^x_{k,i} &= \chi^{x-}_{k,i} + K_k(y_k - \gamma_{k,i}) \\
\hat{x}_k &= \sum_{i=0}^{2n} W^x_i \cdot \chi^x_{k,i} \\
P_{x_k} &= \sum_{i=0}^{2n} W^c_i (\chi^x_{k,i} - \hat{x}_k)(\chi^x_{k,i} - \hat{x}_k)^T
\end{aligned}
$$

Although we have used the fully augmented UKF approach in this algorithm, it can be applied to all the other algorithms previously mentioned by proper adoption of the steps $\chi^x_{k,i}$, $\hat{x}_k$ and P_{x_k} as defined in Table 5.8. Note that if the reformulated correction steps are applied to the algorithm given in Table 5.4, P_{x_k} is guaranteed to be positive semi-definit with α, β, κ as in (5.20), since $W^c_i > 0$. This may not be the case in the other algorithms, since W^c_0 may be negative. Further, P_{x_k} may not be positive semi-definit if the standard calculation $P_{x_k} = P^-_{x_k} - K_k P_{y_k y_k} K^T_k$ is used due to round off errors. Therefore the reformulated correction steps may be preferred in any of the UKF algorithms due to better numerical behavior, but at the expense of computational load.

5.6 Constrained State Estimation

State estimation may introduce some challenges in systems based on first principles. In for example chemical processes one may wish to estimate concentrations. Mathematically the model may allow negative concentration, whilst this is not physically possible. A first principle model is also often nonlinear, and theory based on Gaussian noise may not be applicable, since Gaussian noise propagated through a nonlinear model is no longer Gaussian. Further, nonlinearities may be severe such that theory based on linearizations are too inaccurate and provide poor estimation accuracy. At last the nonlinear system may have a skew or multimodal probability density function (pdf).

One way to interpret the Kalman filter (Simon (2006)) is that it solves the maximum probabilities problem

$$\hat{x}_k = \arg\max_{x_k} pdf(x_k|(y_1^T \cdots y_k^T)) \tag{5.26}$$

given x_0, w_k and v_k Gaussian, i.e. $\hat{x}_k$ is the value of x_k that maximizes $pdf(x_k|(y_1^T \cdots y_k^T))$. Then one may ask what happens if one uses the Kalman filter on a system with a multimodal pdf? A multimodal pdf is illustrated in Figure 5.1. Intuitively, applying any Kalman filter on a system illustrated in Figure 5.1, given (5.26) is true, there may exist two solutions, one for $\hat{x}_k < 0$ and one for $\hat{x}_k > 0$.

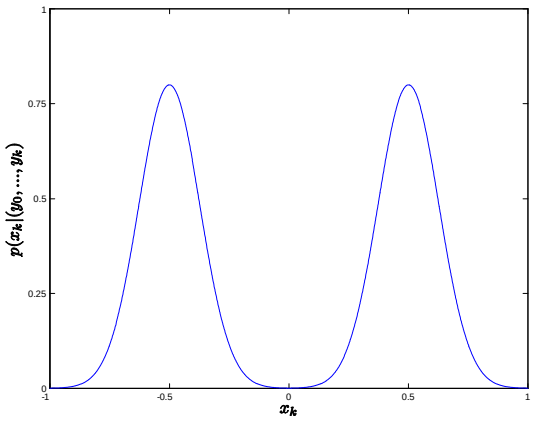

Figure 5.1: An illustration of a multimodal probability density function (pdf)

If $\hat{x}_k$ is an estimate of concentrations in a physical system, we may introduce constraints in order to force the estimates to have physical meaning. Hence the use of constraints may be important.

As mentioned earlier, the literature addressing constraints in the UKF approach is rather limited. In Julier and Uhlmann (1994) the idea of constraining the sigma points is in fact listed as a general possibility, but is however not further discussed. In the following chapters we will turn our focus to constraint implementation.

5.6.1 Heuristic constraint handling

Constraint methods

A common method for implementing constraints in the KF and EKF algorithms is known as clipping, where the corrected state estimate $\hat{x}_k$ is set equal to some predefined bounds d_k if outside these limits (Hasseltine and Rawlings (2003), Simon (2006), Kandepu et al. (2008)).

In Simon and Simon (2005), Simon and Simon (2006) and in Simon (2006) projection methods for Kalman filtering are studied. The unconstrained state estimates $\hat{x}_k$ is projected onto some constrained set $\tilde{x}_k$ by solving the following convex QP problem

$$\min_{\tilde{x}_k}(\tilde{x}_k - \hat{x}_k)^T W_k(\tilde{x}_k - \hat{x}_k) \text{ s.t. } D_k\tilde{x}_k \leq d_k \tag{5.27}$$

(5.27) can be rewritten as the QP problem

$$\min_{\tilde{x}_k}({\tilde{x}_k}^T W_k\tilde{x}_k - 2\hat{x}_k^T W\tilde{x}_k) \text{ s.t } D_k\tilde{x}_k \leq d_k \tag{5.28}$$

By choosing $W_k = I$ we obtain a least squares method, and by choosing $W_k = (P_{x_k})^{-1}$ we get a maximum likelihood estimates (Simon and Simon (2005)). What is worth mentioning is that if $W_k = I$ and $D_k = I$, which correspond to a constraint on each $\hat{x}_k$, the solution of (5.27) is ${\tilde{x}_k}^T = d_k$ in the case when $\hat{x}_k$ violates the constraints d_k(see also Simon (2006)). In this case solving (5.28) gives the same solution as clipping, and we may conclude that clipping is optimal in such cases[7].

Other literature discussing constraint implementation in the UKF approach, apart from Vachhani et al. (2006) and Kandepu et al. (2008) is found in Li and Leung (2004) where they used equality constraints on the

[7] Considering implementation aspects this may have some implications, since the clipping may produce lower computationally load than solving a QP-problem.

corrected estimate, and Julier and Laviola (2007) where two methods for nonlinear equality constraints are proposed.

Constraint candidates

Based on the algorithm steps we are able to identify constraint candidates as given by Table 5.9.

Table 5.9: The constraint candidates

χ_{k-1}	CC_1
$\chi^{x-}_{k,i}$	CC_2
$\hat{x}^-_k$	CC_3
$P^-_{x_k}$	-
$\gamma_{k,i}$	$CC_{4,5}$
$\hat{y}_k$	CC_6
$P_{y_k y_k}$	-
$P_{x_k y_k}$	-
K_k	-
$\chi^x_{k,i}$	CC_7
$\hat{x}_k$	CC_8
P_{x_k}	-

By constraint candidates (CC) we think of steps in the algorithm which may be suitable for some constraint handling, e.g. constraining CC_1 means that χ_{k-1} is constrained according to some set. Kandepu et al. (2008) proposed the constraint candidates CC_1, CC_2, CC_3, CC_5, CC_6 and CC_8 using the UKF algorithm as given in Table 5.7. Further they showed the effect of using the constraint candidate CC_1 on a CSTR case.

By using the algorithm as in Table 5.4 we have the opportunity to extend the list of constraint candidates, by constraining the updated sigma points before they are propagated through the output function. This is indicated by constraint candidate CC_4.

The proposed UKF-formulation of the correction steps as given in Table 5.8 provides the possibility to further extend the list of constraint candidates by constraining each $\chi^x_{k,i}$ before the corrected estimate $\hat{x}_k$ with its associated covariance P_{x_k} is calculated. This is indicated by the constraint candidate CC_7. It is worth noticing that constraining the state estimate $\hat{x}_k$ (CC_8) has no direct impact on the covariance P_{x_k} in all the presented algorithms, except for the NLP UKF, QP UKF (see Chapter 5.2) and the reformulated UKF where the constrained estimate will have a direct impact

on the associated covariance. This aspect, that the constraints is propagated directly through the succeeding covariance calculations, is in our view one of the most important aspects of using the proposed reformulated UKF-algorithm.

Constraints implementation

A logical choice when implementing constraints on the sigma points (CC_1), the propagated sigma points (CC_2) and the updated sigma points (CC_4 and CC_7), is to apply clipping after the calculation step, since no covariance is associated with these constraint candidates. Clipping (or minimum square if there is a algebraic relationship between the constraints) could also be applied to the other constraint candidates (CC_3, CC_6 and CC_8). Alternatively, by instead utilizing the covariance in the constraints calculations, the constrained $\hat{x}_k^-$, $\hat{y}_k$ and $\hat{x}_k$ can be calculated by the maximum likelihood approach as

$$\min_{\tilde{x}_k^-}(\tilde{x}_k^- - \hat{x}_k^-)^T(P_{x_k}^-)^{-1}(\tilde{x}_k^- - \hat{x}_k^-) \text{ s.t. } D_k\tilde{x}_k^- \leq d_{\hat{x}_k^-} \tag{5.29}$$

$$\min_{\tilde{y}_k}(\tilde{y}_k - \hat{y}_k)^T P_{y_k y_k}^{-1}(\tilde{y}_k - \hat{y}_k) \text{ s.t. } D_{\tilde{y}_k}\tilde{y}_k \leq d_{\hat{y}_k} \tag{5.30}$$

$$\min_{\tilde{x}_k}(\tilde{x}_k - \hat{x}_k)^T P_{x_k}^{-1}(\tilde{x}_k - \hat{x}_k) s.t. D_k\tilde{x}_k \leq d_{\hat{x}_k} \tag{5.31}$$

The choice of which method to use, least squares or maximum likelihood, depends on the system and available computational resources.

5.6.2 The NLP/QP-UKF

The Kalman-approach assumes a linear correction of the predicted estimate, see (5.5). In Vachhani et al. (2006) they use the UKF formulation in Julier et al. (1995), i.e. the algorithm as by Table 5.5 with $\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix}$ as in (5.19), and propose a nonlinear correction by solving an NLP. A special property of their NLP algorithm is a recalculation of the weights W if the sigma points χ_{k-1} violates some constraints.

In this work we propose an NLP for the correction step as in Vachhani et al. (2006), but without recalculating the weights W. Further we present the NLP UKF for the fully augmented scaled UKF. Assuming the nonlinear system given by (5.1) - (5.2) corrupted by Gaussian noise, the algorithm is given as in Table 5.10.

Table 5.10: The fully augmented NLP/QP UKF algorithm

$$
\begin{aligned}
\chi_{k-1} &= \begin{bmatrix} \hat{x}_{a_{k-1}} & \hat{x}_{a_{k-1}} + \Gamma\sqrt{P_a} \ldots \hat{x}_{a_{k-1}} - \Gamma\sqrt{P_a} \end{bmatrix} \\
\chi^{x-}_{k,i} &= f(\chi^{x}_{k-1,i}, u_{k-1}, \chi^{v}_{k-1,i}) \\
\hat{x}^{-}_{k} &= \sum_{i=0}^{2n} W^{x}_{i} \cdot \chi^{x-}_{k,i} \\
P^{-}_{x_k} &= \sum_{i=0}^{2n} W^{c}_{i} (\chi^{x-}_{k,i} - \hat{x}^{-}_{k})(\chi^{x-}_{k,i} - \hat{x}^{-}_{k})^{T} \\
\gamma_{k,i} &= h(\chi^{x-}_{k,i}, \chi^{w}_{k,i}) \\
\hat{y}^{-}_{k} &= \sum_{i=0}^{2n} W^{x}_{i} \cdot \gamma_{k,i} \\
P_{y_k y_k} &= - \\
P_{x_k y_k} &= - \\
K_k &= - \\
\chi^{x}_{k,i} &\quad \min_{\chi^{x}_{k,i}} J \\
\hat{x}_k &= \sum_{i=0}^{2n} W^{x}_{i} \cdot \chi^{x}_{k,i} \\
P_{x_k} &= \sum_{i=0}^{2n} W^{c}_{i} (\chi^{x}_{k,i} - \hat{x}_k)(\chi^{x}_{k,i} - \hat{x}_k)^{T}
\end{aligned}
$$

By choosing

$$J = (y_k - h(\chi^x_{k,i}))^T R_k^{-1}(y_k - h(\chi^x_{k,i})) + (\chi^x_{k,i} - \chi^{x-}_{k,i})^T \left(P^-_{x_k}\right)^{-1} (\chi^x_{k,i} - \chi^{x-}_{k,i}) \tag{5.32}$$

as in Vachhani et al. (2006), this leads to a NLP problem. However, by assuming a linear output model, as by (5.4) it can be shown, see Appendix A.3, that the problem reduces to a QP-problem. In this case J becomes

$$J = \chi^{x^T}_{k,i} \left(D_k^T R_k^{-1} D_k + (P_k^-)^{-1}\right) \chi^x_{k,i} - 2(y_k^T R_k^{-1} D_k + \chi^{x-^T}_{k,i} (P_k^-)^{-1}) \chi^x_{k,i} \tag{5.33}$$

Intuitively, respectively, (5.32) and (5.33) balances the correction between the output y_k and the propagated state sigma points $\chi^{x-}_{k,i}$ according to the output uncertainty R_k and the predicted covariance $P^-_{x_k}$. Hereafter we will refer to the algorithm using (5.32) as the NLP UKF and the algorithm using (5.33) as the QP UKF. The QP UKF formulation drastically reduces the computational load compared to the NLP UKF, see Appendix B, since a QP solver generally is less computationally demanding then an NLP solver.

Both the NLP UKF and the QP UKF minimize J with respect to $\chi^x_{k,i}$ such that some constraints are fulfilled. The constraints can typically be given as by (5.34)

$$\begin{gathered} \min_{\chi^x_{k,i}} J \\ st. \\ x_{i_{low}} \leq \chi^x_{k,i} \leq x_{i_{upp}} \\ H_i \chi^x_{k,i} \leq b_i \end{gathered} \tag{5.34}$$

Note that $\chi^x_{k,i}$ can be calculated directly after computing $P^-_{x_k}$. The model output sigma point set $\gamma_{k,i}$ can then be calculated utilizing the updated sigma points $\chi^x_{k,i}$, i.e $\gamma_{k,i} = h(\chi^x_{k,i}, \chi^w_{k,i})$. Note also that the steps regarding $\chi^x_{k,i}$, $\hat{x}_k$ and P_{x_k} as defined in Table 5.10 could be applied to all the other UKF algorithms as well, e.g. to the algorithm as in Table 5.4.

5.7 Notes regarding the algorithms

We have so far presented several algorithms based on the Kalman-approach. It may be worth mentioning that all the unconstrained algorithms reduces

to the linear Kalman filter for linear systems with Gaussian noise. Further we have pointed to several ways of implementing constraints in the presented algorithms. Although the UKF algorithms have been presented as independent algorithms, we will emphasize the fact that the conceptual ideas from one algorithm could easily be used in one of the others, i.e. the reformulated steps, the NLP/QP steps and the constraint handling can be applied to any of the presented UKF algorithms in any combination. For example one can combine the UKF algorithm in Table 5.4 with constraints in the prediction steps and NLP/QP in the correction steps to facilitate nonlinear update in this step. In the simulation chapter to come we will show some examples of such combinations.

5.8 Simulation studies

Several authors have investigated the performance of the UKF and compared it with the EKF or other estimation algorithms, see Julier and Uhlmann (1994), Julier et al. (1995), Julier et al. (1997), Zandt (2001), Julier and Uhlmann (2002), Akin et al. (2003), Laviola (2003), Julier and Uhlmann (2004), van der Merwe (2004), Romanenko and Castro (2004), Romanenko et al. (2004), Rawlings and Bakshi (2006), Xiong et al. (2006), Xiong et al. (2007), Pieper (2007) and Kandepu et al. (2008).

In the following we present some of the presented algorithms' ability to handle problems where the EKF suffers from convergence properties, or even fail to converge. We do not consider the KF, since we study nonlinear systems only in this simulation study.

The case selection is based on cases in Hasseltine and Rawlings (2003) used to indicate the performance of the Moving Horizon Estimation (MHE) approach and at the same time show the limitations using the EKF. The key challenge in the selected cases is that they all have a multi-modal pdf, where some solutions are physically valid, and some not.

5.8.1 Case '2 state CSTR'

Case description

Consider the gas-phase, reversible reaction (Hasseltine and Rawlings (2003))

$$2A \xrightarrow{k_r} B \qquad k_r = 0.16 \tag{5.35}$$

with stoichiometric matrix

$$s = \begin{bmatrix} s_1 & s_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \end{bmatrix} \tag{5.36}$$

and reaction rate

$$r = k_r P_A^2 \tag{5.37}$$

The state and output vectors are defined as

$$x = \begin{bmatrix} P_A \\ P_B \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad y = \begin{bmatrix} 1 & 1 \end{bmatrix} x \tag{5.38}$$

where P_A and P_B are the partial pressures. It is assumed that the ideal gas law holds (high temperature, low pressure), and that the reaction occurs in a well-mixed, isothermal batch reactor. From first principles, the continuos model for this system is

$$\dot{x}(t) = f_c(x(t)) = s^T r(t) \tag{5.39}$$

and the output model

$$y(t) = h_c(x(t)) = \begin{bmatrix} 1 & 1 \end{bmatrix} x(t) \tag{5.40}$$

A discrete analytical solution of (5.39) is

$$x_{1,k+1} = \left(\frac{1}{x_{1,k}} - s_1 k_r \Delta t \right)^{-1} = \frac{x_{1,k}}{1 - s_1 k_r \Delta t \; x_{1,k}} = \frac{x_{1,k}}{1 + 2 k_r \Delta t \; x_{1,k}} \tag{5.41}$$

$$x_{2,k+1} = x_{2,k} + \frac{s_2 k_r}{s_1 k_r} \left(x_{1,k+1} - x_{1,k} \right) = x_{2,k} + \frac{k_r \Delta t \; x_{1,k}}{1 + 2 k_r \Delta t \; x_{1,k}} \tag{5.42}$$

where Δt is the integration step length. Further it is assumed that the system experience Gaussian noise both in the states and in the outputs, given by $v_k \sim N(0, Q_k)$ and $w_k \sim N(0, R_k)$. The discrete system becomes

$$x_{k+1} = f(x_k) + v_k \tag{5.43}$$

and the discrete model of (5.40)

$$y_k = h(x_k) + w_k = \begin{bmatrix} 1 & 1 \end{bmatrix} x_k + w_k \tag{5.44}$$

The parameters used for this system are

$$\begin{aligned} \Delta t = t_{k+1} - t_k = 0.1 \qquad & P_0 = \begin{bmatrix} 6^2 & 0 \\ 0 & 6^2 \end{bmatrix} \\ Q_k = \begin{bmatrix} 0.001^2 & 0 \\ 0 & 0.001^2 \end{bmatrix} \qquad & R_k = 0.1^2 \\ x_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad & \hat{x}_0 = \begin{bmatrix} 0.1 \\ 4.5 \end{bmatrix} \end{aligned} \tag{5.45}$$

Note that the initial guess for the states ($\hat{x}_0$), is very poor. This simple example is used by several authors in order to investigate estimator performance, see Hasseltine and Rawlings (2003), Vachhani et al. (2006), Rawlings and Bakshi (2006), and Kandepu et al. (2008). The reason why this problem is interesting is that the estimator may experience a multimodal pdf, which may lead to unphysical estimates.

Simulation results

In the following chapters we investigate some of the previous described algorithms applied on the 2-state CSTR case. Note that all the parameters are as described in the case description above for all the presented algorithms and that the noise sequences are identical in all simulations. Note also that we have used the exact solution to (5.39). By using Euler or Runge-Kutta integration schemes, the results will differ slightly from the results presented below. This is especially the case in the unconstrained cases. However, the main characteristics are maintained whichever integration scheme is used. We have chosen to be true to the source of these examples (Hasseltine and Rawlings (2003)), and have used the same parameters to achieve comparable results.

EKF Figure 5.2 shows the results of the simulation using unconstrained EKF as by Table 5.3, with numerically derived Jacobians.

As Figure 5.2 shows, the unconstrained EKF fails to converge to the true states within the given time frame[8]. These results are in agreement with the results of Hasseltine and Rawlings (2003) and Kandepu et al. (2008). The reason why the EKF fails is that while the negative pressure is unphysical, the unconstrained estimator allows the estimate to enter regions where the partial pressure may be negative. The behavior of the estimator in this case is best explained "*as a poor initial guess leading to an errant region of attraction*" (Hasseltine and Rawlings (2003)). The "*errant region of attraction*" exist due to the multimodal properties of this system's pdf, see Hasseltine and Rawlings (2003) for a broader discussion of why the EKF fails.

Figure 5.3 shows the results of the simulation using EKF with constraints. The constraint is implemented as clipping the corrected state estimate (CC_8), and the constraints are such that $\hat{x}_k \geq 0$.

[8] Actually the EKF will converge very slovly, but one need to run the simulation approx. 1000 samples.

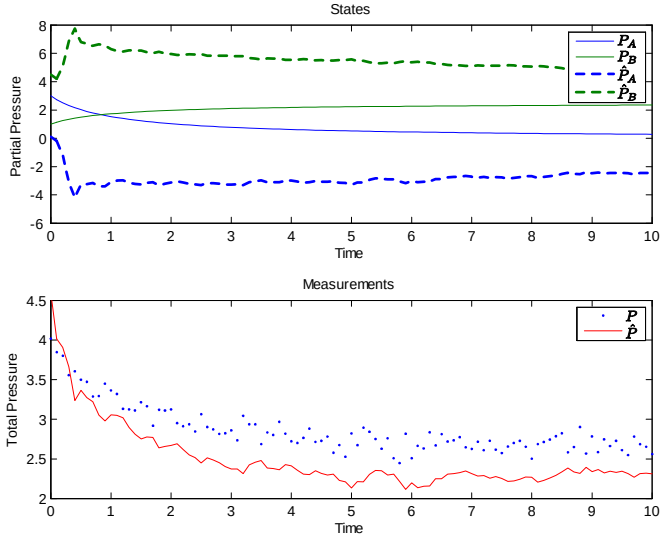

Figure 5.2: Unconstrained EKF

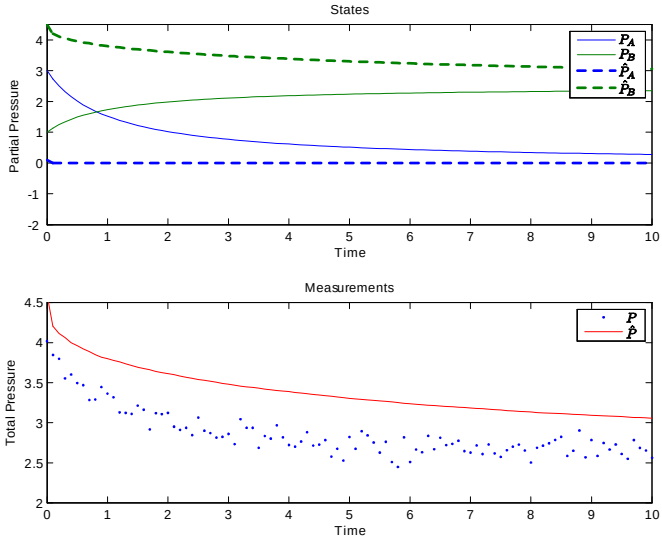

Figure 5.3: Constrained EKF

As Figure 5.3 shows, the constrained EKF fails to converge to the true states. These results are in agreement with the results of Hasseltine and Rawlings (2003) and Kandepu et al. (2008). By clipping the state $\hat{x}_k$, it is restricted to a valid physical region, but the knowledge about the constraints is not propagated into the covariance, and the hence accuracy of the approximated covariance matrix P_{x_k} is questionable. As Figure 5.3 indicates, the performance of the clipped EKF is rather poor.

UKF In the following chapters a selection of the presented Jacobian free algorithms is tested on the 2 state CSTR case. For the UKF algorithms using (5.12) for sigma point selection, (5.18) is used for α, β and κ. For the UKF algorithms using (5.11) for the sigma point selection, the values given by (5.20) is used for α, β and κ.

First we investigate the unconstrained fully augmented UKF given by Table 5.7 and using the Cholesky square root algorithm. Figure 5.4 shows the results of the simulation.

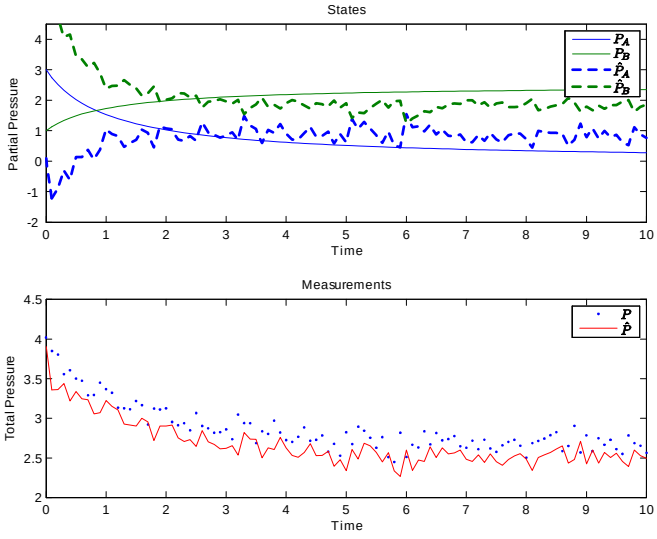

Figure 5.4: Unconstrained fully augmented UKF

As Figure 5.4 shows, this algorithm fails to converge to the true states within the given time frame. These results are in agreement with the results of Kandepu et al. (2008). Assuming (5.26) also is valid for the UKF, the unconstrained UKF as well as the EKF has to deal with multiple optima,

and we believe this is the reason why the unconstrained UKF algorithms suffer poor performance on this case.

Next, we extend the fully augmented UKF as in Table 5.7 with constraint handling and Cholesky square root. The constraints is such that $\chi^x_{k-1} \geq 0$ (CC_1). Figure 5.5 shows the results of the simulation.

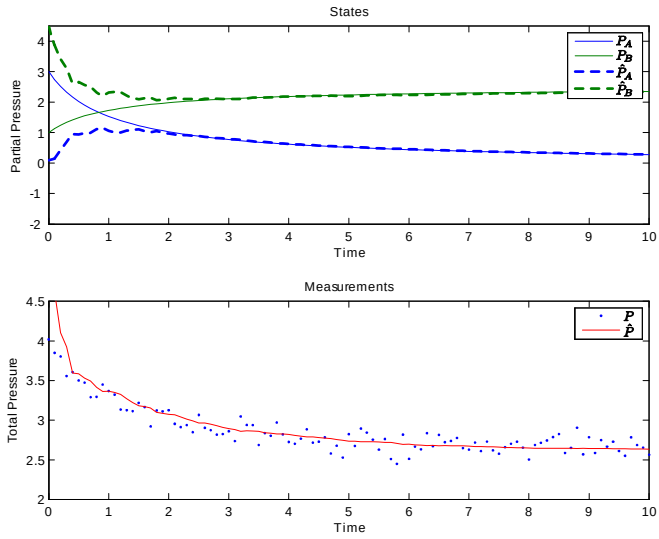

Figure 5.5: Constrained ($\chi^x_{k-1} \geq 0$ (CC_1)) fully augmented UKF

As Figure 5.5 shows, this algorithm converges to the true states after approximately 25 samples regards ($t = 2.5s$). These results are in agreement with the results achieved by Kandepu et al. (2008). By constraining $\chi^x_{k-1} \geq 0$ (CC_1) the sigma points are restricted to physical valid values. Also, by inspecting the structure of the UKF algorithm, one sees (e.g. see Table 5.7) that constraining the sigma points not only propagates the constraints to the mean value, but also influences the covariance calculations. By this the covariance is related to the mean value, and the unphysical region of attraction is limited. Note that since the sigma points are propagated through the model after they are constrained, constraining the sigma points only (CC_1) does not generally guarantee that the state estimates does not enter an unphysical value.

All the UKF algorithms converge to the true state when constraining the sigma points $\chi^x_{k-1} \geq 0$ (CC_1). Although the unconstrained UKF in Table 5.4 with symmetric square root calculation may converge, it is possible to

increase the convergence performance. By using the reformulated correction steps from the algorithm in Table 5.8 and constraint handling by using CC_1 ($\chi^x_{k-1} \geq 0$) and CC_7 ($\chi^x_k \geq 0$) applied on the non-augmented algorithm in Table 5.4, we get the results as given in Figure 5.6.

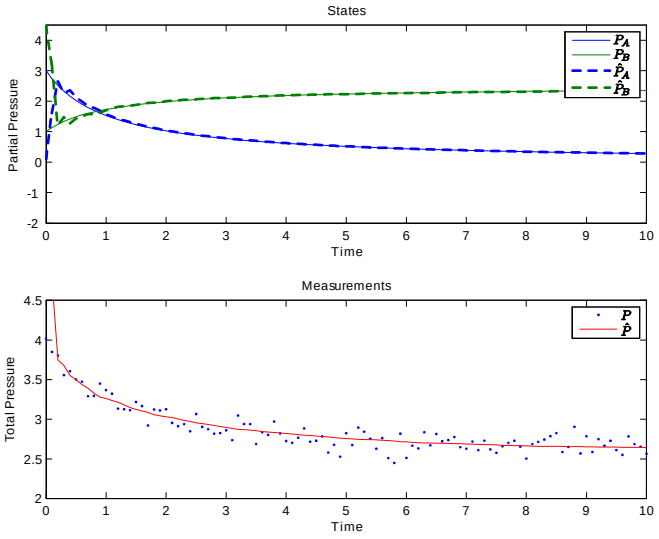

Figure 5.6: Constrained (CC_1 and CC_7) non-augmented UKF as by D.Simon

In fact, the combination of constraint handling and the reformulated steps makes this algorithm to almost converge after only 2 samples ($0.2s$) for this particular noise sequence.

Next, we investigate the QP UKF by combining Table 5.10 and (5.33) with the algorithm presented in Table 5.4, again using the symmetric square root formulation.

The system under investigation has a linear output model, and hence the results of the QP UKF are identical to the NLP UKF (Table 5.10). The results of the simulation are shown in Figure 5.7.

As Figure 5.7 shows, the NLP/QP-algorithm converge after 2 samples ($0.2s$). Notice also the lack of the initial spike in the output. This is because the output in this situation is selected using the optimized sigma point set $\chi^x_{k,i}$ instead of $\chi^{x-}_{k,i}$.

Although the QP UKF and NLP UKF produces the same estimate in the case of a linear output model, there is a huge difference when it comes

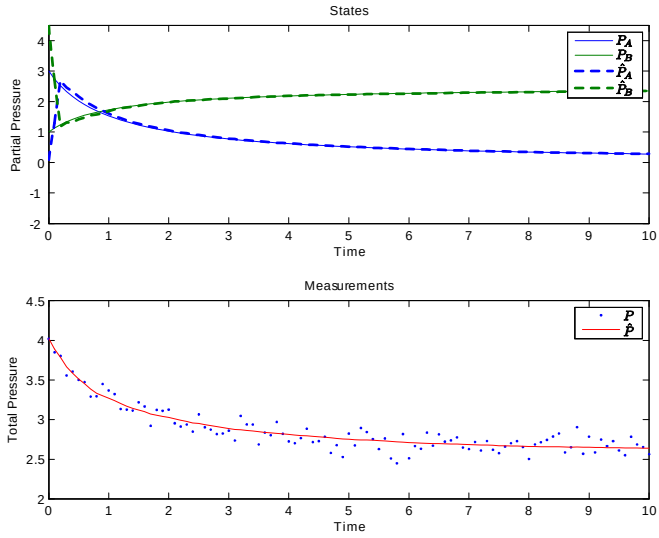

Figure 5.7: QP-UKF using symmetric square root calculations

to computational load. Used on the 2 state CSTR problem, the QP UKF is approximately 8 times faster than the NLP UKF[9] when using the algorithm in Table 5.4 as a basis, see Appendix B. Using the QP UKF described in Table 5.10 is approximately 11 times faster than the NLP UKF. In this case we also experience best convergence performance using the symmetric square root and not the Cholesky square root algorithm.

[9]Implemented in Matlab using the quadprog() for the QP problem and fmincon() for the NLP problem.

5.8.2 Case '3 State Batch reactor'

Case description

Consider the gas-phase, reversible reaction (Hasseltine and Rawlings (2003))

$$A \underset{k_2}{\overset{k_1}{\leftrightarrows}} B + C \tag{5.46}$$

$$2B \underset{k_4}{\overset{k_3}{\leftrightarrows}} C \tag{5.47}$$

$$k = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.05 & 0.2 & 0.01 \end{bmatrix} \tag{5.48}$$

with stoichiometric matrix

$$v = \begin{bmatrix} -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix} \tag{5.49}$$

and reaction rate

$$r = \begin{bmatrix} k_1 c_A - k_2 c_B c_C \\ k_3 c_B^2 - k_4 c_C \end{bmatrix} \tag{5.50}$$

The state and output vectors are defined as

$$x = \begin{bmatrix} c_A \\ c_B \\ c_C \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad y = \begin{bmatrix} RT & RT & RT \end{bmatrix} x, \quad RT = 32.84 \tag{5.51}$$

where c_i is the concentration of the species. It is assumed that the ideal gas law holds, and that the reaction occurs in a well-mixed, constant volume, isothermal batch reactor. From first principles, the continuos model for this system is

$$\dot{x}(t) = f_c(x(t)) = v^T r(t) \tag{5.52}$$

and the output (measurement) model

$$y(t) = h(x(t)) = \begin{bmatrix} RT & RT & RT \end{bmatrix} x(t) \tag{5.53}$$

The model is discretized by using Runge-Kutta 4^{th} order method. Further it is assumed that the system experiences Gaussian noise both in the state and in the measurements, respectively $w_k \sim N(0, Q_k)$ and $v_k \sim N(0, R_k)$. The discrete system is given by

$$x_{k+1} = f(x_k) + w_k \tag{5.54}$$

$$y_k = h(x_k) + v_k = \begin{bmatrix} RT & RT & RT \end{bmatrix} x_k + v_k \tag{5.55}$$

The parameters used for this system are

$$\Delta t = t_{k+1} - t_k = 0.25 \qquad P_0 = \begin{bmatrix} 0.5^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 0.5^2 \end{bmatrix}$$
$$Q_k = \begin{bmatrix} 0.001^2 & 0 & 0 \\ 0 & 0.001^2 & 0 \\ 0 & 0 & 0.001^2 \end{bmatrix} \qquad R_k = 0.25^2 \tag{5.56}$$
$$x_0 = \begin{bmatrix} 0.5 \\ 0.05 \\ 0 \end{bmatrix} \qquad \hat{x}_0 = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$$

It may be noted that the initial guess for the estimator ($\hat{x}_0$), is very poor. Again the estimator may encounter a multi-modal pdf, which may lead to unphysical estimates.

Simulation results

In the next chapters we present some results of our investigation on the 3-state CSTR case. Note that all the parameters are as described in the case description above for all the presented algorithms. Also the noise sequences w_k and v_k are identical in all simulations. The performance of the EKF and the constrained (clipped) EKF used on this case is well documented in Hasseltine and Rawlings (2003), and since our results is in full agreement with those results regarding the EKF, we have chosen not to include them here.

Further, none of the unconstrained UKF algorithms presented in this work converged within the time limit. Hence, the following results are focused on the constrained approach.

The fully augmented UKF as in Table 5.7 is first applied, using the Cholesky square root algorithm, constrained sigma points (CC_1) and $\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix}$ as in (5.18).

As Figure 5.8 shows, the performance is rather poor when it comes convergence speed. Note that all of the constrained UKF algorithms converge within the time limit, but uses unacceptable long time when constraining only the sigma point set by $\chi^x_{k-1} \geq 0$ (CC_1).

Second, we investigate the reformulated fully augmented UKF (Table 5.8) with the Cholesky square root algorithm and constrained sigma points (CC_1 and CC_7), constrained predicted state estimates (CC_3) and

$$\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix} = \begin{bmatrix} 0.7 & 0 & 3-n \end{bmatrix} \tag{5.57}$$

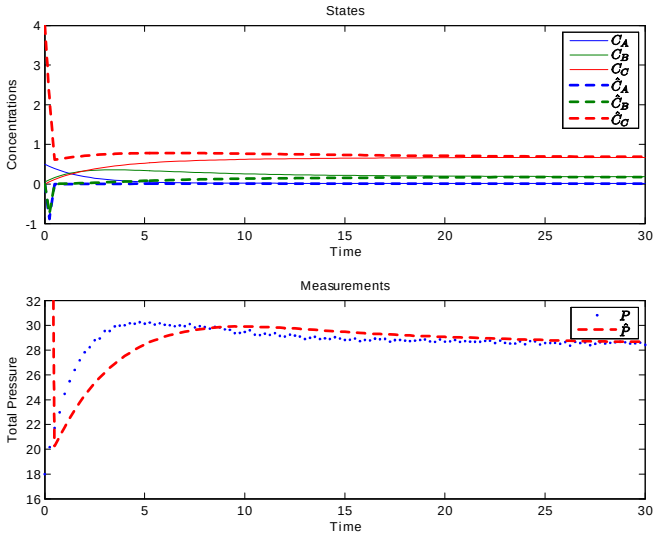

Figure 5.8: The figure shows the performance of the UKF algorithm based on the algorithm in Table 5.7 with constraint handling using CC_1.

By using the constraints

$$\begin{aligned} \chi^x_{k-1} &\geq 0 \ (CC_1) \\ \chi^x_{k,i} &\geq 0 \ (CC_7) \\ x^-_k &\leq \begin{bmatrix} \infty & \infty & 4 \end{bmatrix}^T \ (CC_3) \end{aligned} \tag{5.58}$$

we got the results as shown in Figure 5.9. Convergence to the correct states is fast.

The results of the other UKF algorithms, using the reformulated correction steps, the same $\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix}$-set as defined by (5.57)[10] and constraints as defined in (5.58), have all similar performance.

Third, we study the NLP/QP UKF with additive system noise, i.e. the algorithm in Table 5.5 with $\chi^x_{k,i}$, $\hat{x}_k$ and P_{x_k} as in Table 5.10. Further, the symmetric square root algorithm is used, constrained sigma points (CC_1), constrained predicted state estimate x^-_k (CC_3) and

$$\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix} = \begin{bmatrix} 1 & 10 & 0 \end{bmatrix} \tag{5.59}$$

[10] Note that the algorithms using eq.(5.11) for the sigma point selection, uses (5.20) as their $\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix}$-set.

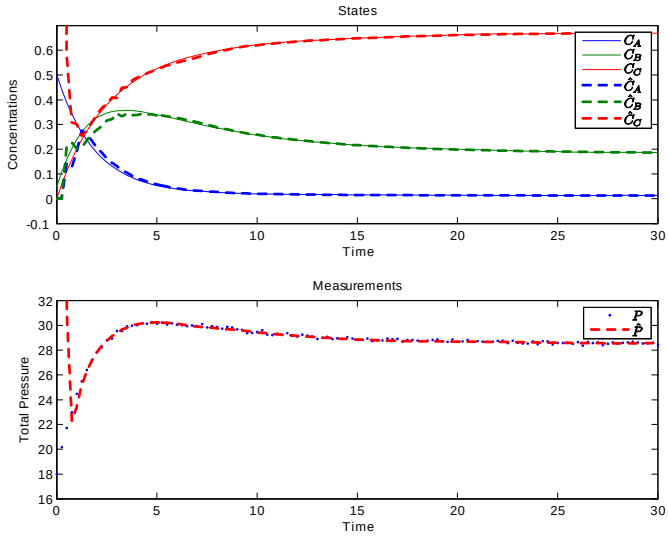

Figure 5.9: The figure shows the performance of the UKF algorithm based on the algorithm in Table(5.8) with constraints. The constraints used are CC_1, CC_3 and CC_7.

By using the constraints

$$\chi^x_{k-1} \geq 0 \quad (CC_1) \tag{5.60}$$

$$x^-_k \leq \begin{bmatrix} \infty & \infty & 4 \end{bmatrix}^T (CC_3) \tag{5.61}$$

and minimize $\chi^x_{k,i}$ such that

$$0 \leq \chi^x_{k,i} \leq \begin{bmatrix} \infty & \infty & 4 \end{bmatrix}^T \quad (\text{i.e.} CC_7 \text{ by QP}) \tag{5.62}$$

we got the results as shown in Figure 5.10.

As Figure 5.10 shows, the performance is further improved. Note that the use of the constraints (CC_1) and (CC_3) was required in order to achieve these results, as well as some tuning of the α, β and κ (for a broader general discussion of the effect of α, β and κ see Julier and Uhlmann (2002), Julier and Uhlmann (2004)).

5.9 Discussion

We have shown several UKF based algorithms and proposed some methods to implement constraints using these algorithms. Further we have indicated

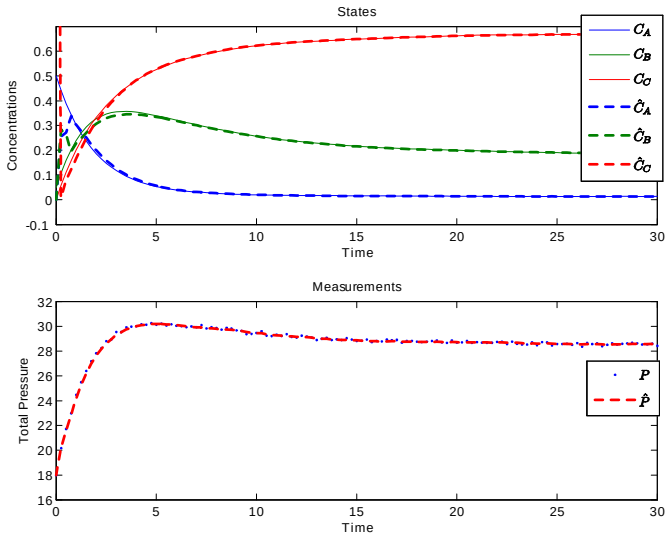

Figure 5.10: The figure shows the performance of the QP UKF algorithm based on the algorithm in Table 5.5 with the correction steps as in Table 5.10. The constraints used are CC_1, CC_3 and CC_7.

that the convergence performance of the algorithms is not only different, but for the same algorithm might be influenced by the selection of the square root algorithm[11], at which algorithm step the constraints are placed, and the use of the reformulated correction steps. Figure 5.11 further illustrates this.

As Figure 5.11 demonstrates there is an effect of using both the reformulated correction steps and the symmetric square root calculation in the 2-state CSTR case with respect to convergence speed.

Compared to the results in Kandepu et al. (2008) and the results presented in Figure 5.5, we have shown that it is possible to decrease the convergence time from 25 ($2.5s$) samples to approximately 2 samples ($0.2s$) (see Figure 5.11 d). Compared to the unconstrained EKF, which converges in approximately 1000 samples with this particular noise sequence, this indicates the superior performance of the suggested constrained UKF-approach applied to the 2-state CSTR case[12].

Also, by applying UKF to the 3-state case, we have demonstrated that

[11] See Appendix B.

[12] See Appendix B for performance measurements regard the 'Case 2A->B'.

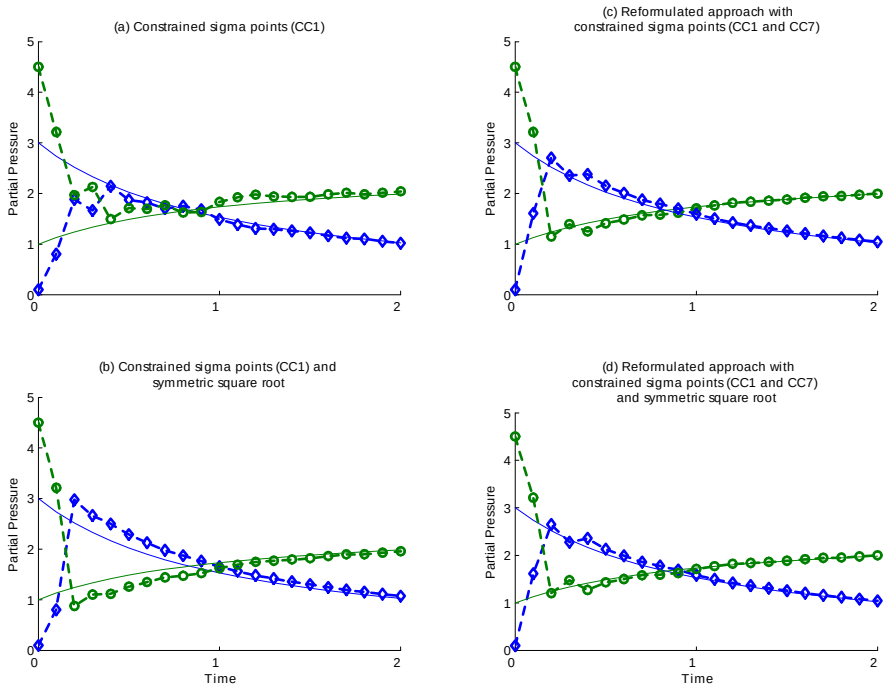

Figure 5.11: A comparison of the effect of the different formulation alternatives. The figure shows four sub-figures with the states of the 20 first samples of the 2-state CSTR case, using the non-augmented UKF as in Table 5.4, weights (W) by (5.17) and constraining $\chi^x_{k-1} \geq 0$ (CC_1). In (a) the Cholesky square root algorithm is used. (b) shows the change in the estimates from (a) as the Cholesky square root algorithm is replaced with a symmetric square root. (c) shows the change in the estimates from sub-figure (a) as the reformulated correction steps is introduced along with constraining $\chi^x_{k,i} \geq 0$. (d) shows the change in the estimates from sub-figure (a) as the Cholesky square root is replaced with a symmetric square root and the reformulated correction steps is introduced constraining $\chi^x_{k,i} \geq 0$ (CC_7).

the constrained reformulated approach achieves good performance. The constrained UKF, constraining only $\chi^x_{k-1} \geq 0$ (CC_1), suffered from poor performance with respect to convergence speed, and converged after approximately 70 samples. The QP UKF converged after approximately 3 samples, and the reformulated fully augmented constrained UKF after approximately 5 samples.

However, care should be taken. Applying constraints as suggested in this paper may introduce singularities and even indefinit covariance matrices. This may happen e.g. if all the constrained sigma points of one state are outside the constraints at the same time as the estimated state also is outside the constraint. Strategies to identify and overcome such situations may be required. Several strategies could be applied, but it is beyond the scope (and space) of this paper to enter that discussion.

By introducing both a reformulation of the correction steps and the use of constraints in the UKF approach, the UKF may not be considered 'unscented' any more. To distinguish from the original UKF, it is maybe more correct to refer to the versions utilizing constraints and/or reformulated corrections steps as the Constrained Unscented Kalman filter (CUKF).

Further, Table 5.11-5.13 in Appendix B indicates that adding complexity to the original UKF algorithms have a computational cost, but as Table 5.14-5.15 in the appendix indicates, with a potential increase in estimation accuracy, even though the most complex not necessarily gives the most accurate estimation. Given a system, which algorithm to select is a trial and error process.

5.10 Conclusions

We have in this paper given a broad overview of several UKF based nonlinear estimation algorithms as an alternative to the EKF, suggested a reformulation of the correction step which can be applied to all of the presented UKF algorithms, presented a QP formulation of the NLP UKF (which also can be applied to all of the presented UKF algorithms) and proposed alternatives to realize constraints within the UKF approach.

Further we have demonstrated the superior performance of the constrained UKF approach over the EKF approach applied to both a '2-state CSTR' example system and a '3-state batch' example system.

5.11 Acknowledgements

The financial support of Hydro Aluminium AS is gratefully acknowledged.

5.12 Appendix A - Algebraic proofs

5.12.1 Proof of reformulated corrected state estimate

We want to show that

$$\hat{x}_k = \sum_{i=0}^{2n} W_i^x \cdot \chi_{k,i}^x = \hat{x}_k^- + K_k(y_k - \hat{y}_k^-) \tag{5.63}$$

Define

$$\chi_{k,i}^x = \chi_{k,i}^{x-} + K_k(y_k - h(\chi_{k,i}^{x-}\chi_{k,i}^w)) \tag{5.64}$$

then

$$\begin{aligned}
\hat{x}_k = \sum_{i=0}^{2n} W_i^x \cdot \chi_{k,i}^x &= \sum_{i=0}^{2n} W_i^x \left(\chi_{k,i}^{x-} + K_k(y_k - h(\chi_{k,i}^{x-}\chi_{k,i}^w))\right) \\
&= \sum_{i=0}^{2n} W_i^x \chi_{k,i}^{x-} + \sum_{i=0}^{2n} W_i^x K_k(y_k - h(\chi_{k,i}^{x-}\chi_{k,i}^w)) \\
&= \sum_{i=0}^{2n} W_i^x \chi_{k,i}^{x-} + K_k \sum_{i=0}^{2n} W_i^x (y_k - h(\chi_{k,i}^{x-}\chi_{k,i}^w)) \\
= \sum_{i=0}^{2n} W_i^x \chi_{k,i}^{x-} + K_k &\left(\sum_{i=0}^{2n} W_i^{x-} y_k - \sum_{i=0}^{2n} W_i^x h(\chi_{k,i}^{x-}\chi_{k,i}^w)\right) \\
&= \hat{x}_k^- + K_k(y_k - \hat{y}_k)
\end{aligned} \tag{5.65}$$

Note $\sum_{i=0}^{2n} W_i^x = 1$ and the other definitions as given by Table 5.8 for the reformulated UKF.

5.12.2 Proof of the reformulated corrected covariance

We want to show that

$$P_{x_k} = \sum_{i=0}^{2n} W_i^c(\chi_{k,i}^x - \hat{x}_k)(\chi_{k,i}^x - \hat{x}_k)^T = P_{x_k}^- - K_k P_{y_k y_k} K_k^T \tag{5.66}$$

where

$$P_{x_k}^{-} > 0 \tag{5.67}$$

Define

$$\chi_{k,i}^{x} = \chi_{k,i}^{x-} + K_k(y_k - h(\chi_{k,i}^{x-}\chi_{k,i}^{w})) = \chi_{k,i}^{x-} + K_k(y_k - \gamma_{k,i}) \tag{5.68}$$

then, by using the definitions given in Table 5.8 for the reformulated UKF we obtain

$$
\begin{aligned}
P_{x_k} &= \sum_{i=0}^{2n} W_i^c(\chi_{k,i}^x - \hat{x}_k)(\chi_{k,i}^x - \hat{x}_k)^T \\
&= \sum_{i=0}^{2n} W_i^c(\chi_{k,i}^x - (\hat{x}_k^- + K_k(y_k - \hat{y}_k^-)))(\chi_{k,i}^x - (\hat{x}_k^- + K_k(y_k - \hat{y}_k^-)))^T \\
&= \sum_{i=0}^{2n} W_i^c(\chi_{k,i}^{x^-} + K_k(y_k - \gamma_{k,i}) - \hat{x}_k^- - K_k(y_k - \hat{y}_k^-))\cdot \\
&\quad (\chi_{k,i}^{x^-} + K_k(y_k - \gamma_{k,i}) - \hat{x}_k^- - K_k(y_k - \hat{y}_k^-))^T \\
&= \sum_{i=0}^{2n} W_i^c((\chi_{k,i}^{x^-} - \hat{x}_k^-) + K_k(y_k - y_k - \gamma_{k,i} + \hat{y}_k^-))\cdot \\
&\quad ((\chi_{k,i}^{x^-} - \hat{x}_k^-) + K_k(y_k - y_k - \gamma_{k,i} + \hat{y}_k^-))^T \\
&= \sum_{i=0}^{2n} W_i^c((\chi_{k,i}^{x^-} - \hat{x}_k^-) + K_k(\gamma_{k,i} + \hat{y}_k^-))\cdot \\
&\quad ((\chi_{k,i}^{x^-} - \hat{x}_k^-) + K_k(\gamma_{k,i} + \hat{y}_k^-))^T \\
&= \sum_{i=0}^{2n} W_i^c((\chi_{k,i}^{x^-} - \hat{x}_k^-)(\chi_{k,i}^{x^-} - \hat{x}_k^-)^T - \\
&\quad (\chi_{k,i}^{x^-} - \hat{x}_k^-)(\gamma_{k,i} - \hat{y}_k^-)^T K_k^T - K_k(\gamma_{k,i} - \hat{y}_k^-)(\chi_{k,i}^{x^-} - \hat{x}_k^-)^T + \\
&\quad K_k(\gamma_{k,i} - \hat{y}_k^-)(\gamma_{k,i} - \hat{y}_k^-)^T K_k^T) \\
&= \sum_{i=0}^{2n} W_i^c(\chi_{k,i}^{x^-} - \hat{x}_k^-)(\chi_{k,i}^{x^-} - \hat{x}_k^-)^T - \sum_{i=0}^{2n} W_i^c(\chi_{k,i}^{x^-} - \hat{x}_k^-)(\gamma_{k,i} - \hat{y}_k^-)^T K_k^T \\
&\quad - K_k \sum_{i=0}^{2n} W_i^c(\gamma_{k,i} - \hat{y}_k^-)(\chi_{k,i}^{x^-} - \hat{x}_k^-)^T + K_k \sum_{i=0}^{2n} W_i^c(\gamma_{k,i} - \hat{y}_k^-)(\gamma_{k,i} - \hat{y}_k^-)^T K_k^T \\
&= P_{x_k}^- - P_{x_k y_k} K_k^T - K_k P_{x_k y_k}^T + K_k P_{y_k y_k} K_k^T \\
&= P_{x_k}^- - P_{x_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T - P_{x_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T + P_{x_k y_k} P_{y_k y_k}^{-1} P_{y_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T \\
&= P_{x_k}^- - P_{x_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T - P_{x_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T + P_{y_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T \\
&= P_{x_k}^- - P_{x_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T \\
&= P_{x_k}^- - P_{x_k y_k} P_{y_k y_k}^{-1} P_{y_k y_k} P_{y_k y_k}^{-1} P_{x_k y_k}^T \\
&= P_{x_k}^- - K_k P_{y_k y_k} K_k^T
\end{aligned}
\tag{5.69}
$$

5.12.3 Proof of the QP-formulation

Vachhani et al. (2006) suggested to use an NLP to calculate an updated sigma point set. An adoption of Vachhani's approach leads to the NLP formulation

$$\min_{\chi^{x+}_{k,i}} \left\{ \begin{array}{c} (y_k - h(\chi^x_{k,i}))^T R_k^{-1} (y_k - h(\chi^x_{k,i}))+ \\ (\chi^x_{k,i} - \chi^{x-}_{k,i})^T \left(P^-_{x_k}\right)^{-1} (\chi^x_{k,i} - \chi^{x-}_{k,i}) \end{array} \right\} \tag{5.70}$$

If the output function $h(\cdot)$ is linear we can formulate the problem

$$\begin{aligned} &\min_{\chi^{x+}_{k,i}} \left\{ \begin{array}{c} (y_k - D_k\chi^x_{k,i})^T R_k^{-1} (y_k - D_k\chi^x_{k,i}))+ \\ (\chi^x_{k,i} - \chi^{x-}_{k,i})^T \left(P^-_{x_k}\right)^{-1} (\chi^x_{k,i} - \chi^{x-}_{k,i}) \end{array} \right\} \\ &= \min_{\chi^{x+}_{k,i}} \left\{ \begin{array}{c} \left(y_k^T R_k^{-1} - {\chi^x_{k,i}}^T D_k^T R_k^{-1}\right)\left(y_k - D_k\chi^x_{k,i}\right) + \\ \left({\chi^x_{k,i}}^T \left(P^-_{x_k}\right)^{-1} - {\chi^{x-}_{k,i}}^T \left(P^-_{x_k}\right)^{-1}\right)(\chi^x_{k,i} - \chi^{x-}_{k,i}) \end{array} \right\} \\ &= \min_{\chi^{x+}_{k,i}} \left\{ \begin{array}{c} y_k^T R_k^{-1} y_k - {\chi^x_{k,i}}^T D_k^T R_k^{-1} y_k - y_k^T R_k^{-1} D_k \chi^x_{k,i} + \\ {\chi^x_{k,i}}^T D_k^T R_k^{-1} D_k \chi^x_{k,i} + {\chi^x_{k,i}}^T \left(P^-_{x_k}\right)^{-1} \chi^x_{k,i} - \\ {\chi^{x-}_{k,i}}^T \left(P^-_{x_k}\right)^{-1} \chi^x_{k,i} - {\chi^x_{k,i}}^T \left(P^-_{x_k}\right)^{-1} \chi^{x-}_{k,i} + \\ {\chi^{x-}_{k,i}}^T \left(P^-_{x_k}\right)^{-1} \chi^{x-}_{k,i} \end{array} \right\} \end{aligned} \tag{5.71}$$

Assume symmetric positive definit R_k and P_k and get

$$\min_{\chi^{x+}_{k,i}} \left\{ \begin{array}{c} y_k^T R_k^{-1} y_k - 2y_k^T R_k^{-1} D_k \chi^x_{k,i} + {\chi^x_{k,i}}^T D_k^T R_k^{-1} D_k \chi^x_{k,i} + \\ {\chi^x_{k,i}}^T \left(P^-_{x_k}\right)^{-1} \chi^x_{k,i} - 2{\chi^{x-}_{k,i}}^T \left(P^-_{x_k}\right)^{-1} \chi^x_{k,i} + {\chi^{x-}_{k,i}}^T \left(P^-_{x_k}\right)^{-1} \chi^{x-}_{k,i} \end{array} \right\} \tag{5.72}$$

However minimizing (5.72) is the same as minimizing the QP-problem

$$
\begin{aligned}
&\min_{\chi_{k,i}^{x+}} \left\{ \begin{array}{c} \left({\chi_{k,i}^{x}}^T D_k^T R_k^{-1} D_k \chi_{k,i}^x - 2 y_k^T R_k^{-1} D_k \chi_{k,i}^x \right) + \\ \left({\chi_{k,i}^{x}}^T \left(P_{x_k}^- \right)^{-1} \chi_{k,i}^x - 2 {\chi_{k,i}^{x-}}^T \left(P_{x_k}^- \right)^{-1} \chi_{k,i}^x \right) \end{array} \right\} \\
&= \min_{\chi_{k,i}^{x+}} \left\{ {\chi_{k,i}^{x}}^T \underbrace{\left(D_k^T R_k^{-1} D_k + \left(P_{x_k}^- \right)^{-1} \right)}_{H_k} \chi_{k,i}^x - 2 \underbrace{\left(y_k^T R_k^{-1} D_k + {\chi_{k,i}^{x-}}^T \left(P_{x_k}^- \right)^{-1} \right)}_{f_{k_i}^T} \chi_{k,i}^x \right\} \\
&= \min_{\chi_{k,i}^{x+}} \left\{ {\chi_{k,i}^{x}}^T H_k \chi_{k,i}^x - 2 f_{k_i}^T \chi_{k,i}^x \right\} \\
&= \min_{\chi_{k,i}^{x+}} \left\{ \begin{array}{c} \frac{1}{2} {\chi_{k,i}^{x}}^T \underbrace{2 H_k}_{\tilde{H}_k} \chi_{k,i}^x - \underbrace{2 f_{k_i}^T}_{\tilde{f}_{k_i}^T} \chi_{k,i}^x \end{array} \right\} \\
&= \min_{\chi_{k,i}^{x+}} \left\{ \frac{1}{2} {\chi_{k,i}^{x}}^T \tilde{H}_k \chi_{k,i}^x - \tilde{f}_{k_i}^T \chi_{k,i}^x \right\}
\end{aligned}
\tag{5.73}
$$

5.13 Appendix B - Performance measures

All simulations and development of the algorithms was done in Matlab Release 14. The computer used was a Dell Precision M70 with Intel Pentium M processor running at 2.13Ghz and 2GB RAM.

5.13.1 Computational Performance

As a guide to compare and determine the performance of the various algorithms we measured the average time to run 10 simulations of the case '2 state CSTR'. The results are shown in Table 5.11-5.13. The notation in the tables is as follows: Column labeled Ch: Simulated using Cholesky matrix square root. Column labeled Sy: Simulated using symmetric matrix square root. Column labeled Ch+Re: Simulated using Cholesky matrix square root and the reformulated correction steps. Column labeled Sy+Re: Simulated using symmetric matrix square root and the reformulated correction steps. All values are in seconds [s].

Note that the results in Table 5.11 are performance measures only, since none of the unconstrained algorithms gave acceptable performance with

Table 5.11: The table shows the average time to run 10 simulations of the '2 state CSTR' case using unconstrained versions of the algorithms.

	Ch [s]	Sy [s]	Ch+Re [s]	Sy+Re [s]
EKF (Table 5.3)	0.4	0.4	0.4	0.4
UKF D.Simon (Table 5.4)	0.4	0.5	0.4	0.5
UKF Non-aug (Table 5.5)	0.5	0.6	0.6	0.6
UKF Q-aug (Table 5.6)	0.7	0.8	0.8	0.9
UKF Fully-aug (Table 5.7)	0.8	0.9	0.9	0.9

respect to the estimation accuracy. The results in Table 5.11 indicate that some of the UKF algorithms in its plain version are not necessarily more computational demanding than the EKF, given the tested case. However, as the number of the sigma points increase, the computational demand increases as well.

Table 5.12: The table shows the average time to run 10 simulations of the '2 state CSTR' case using constrained versions of the UKF algorithms.

	Ch [s]	Sy [s]	Ch+Re [s]	Sy+Re [s]
UKF D.Simon (Table 5.4)	0.8	0.9	1.2	1.2
UKF Non-aug (Table 5.5)	1.0	1.0	1.4	1.5
UKF Q-aug (Table 5.6)	1.7	1.6	2.4	2.5
UKF Fully-aug (Table 5.7)	1.9	2.0	2.9	2.9

Note that for the simulation results shown in Table 5.12 the constraints used are $\chi^x_{k-1} \geq 0 \;\; (CC_1)$ in all the algorithms. For the simulations using the reformulated correction steps, the constraints $\chi^x_k \geq 0 (CC_7)$ are also used. The notation in the table is as follows: Column labeled QP Ch: Simulated using Cholesky square root and the QP version of the correction steps. Column labeled QP Sy: Simulated using symmetric square root and the QP version of the correction steps. Column labeled NLP Ch: Simulated using Cholesky square root and the NLP version of the correction steps. Column labeled NLP Sy: Simulated using Symmetric matrix square root and the NLP version of the correction steps. All values are in seconds [s]

Note also that for the simulation results shown in Table 5.13 the constraints used are $\chi^x_{k-1} \geq 0 \;\; (CC_1)$ in all the algorithms. For the QP UKF's the cost function used is as by (5.33). For the NLP UKF's the cost function used is as by (5.32).

Table 5.11-5.13 indicates that adding complexity to the UKF algorithms

Table 5.13: The table shows the average time to run 10 simulations of the '2 state CSTR' case using the QP/NLP UKF algorithms.

	QP Ch [s]	QP Sy [s]	NLP Ch [s]	NLP Sy [s]
UKF D.Simon (Table 5.4)	1.3	1.3	12.2	12.1
UKF Non-aug (Table 5.5)	1.6	1.6	15.1	17.4
UKF Q-aug (Table 5.6)	2.6	2.6	29.6	29.0
UKF Fully-aug (Table 5.7)	3.0	3.1	35.9	38.3

clearly has a computational cost, but as will be shown below in Appendix B, with a potential increase in estimation accuracy.

5.13.2 Estimation Performance

As a guide to compare and determine the performance of the various algorithms we measured the average square root error of the true and estimated state by 100 Monte Carlo simulations of the case '2 state CSTR'. The error was measured as

$$\bar{e} = \frac{1}{100}\sum_{j=1}^{100}(\tilde{e}_{i,j})\,; \quad i \in [1,2] \tag{5.74}$$

$$\tilde{e}_{i,j} = \frac{1}{100}\sum_{k=1}^{100}\sqrt{(x_{i,k}-\hat{x}_{i,k})\,(x_{i,k}-\hat{x}_{i,k})^T}; \quad i \in [1,2] \tag{5.75}$$

where $\tilde{e}$ is a vector for the average square root error in simulation run j, and $\bar{e}$ is a vector containing the average value of the states of all the average square root errors.

The results are shown in Table 5.14-5.15[13] . The notation in the table is as follows: Column labeled Ch: Simulated using Cholesky square root. Column labeled Sy: Simulated using symmetric square root. Column labeled Ch+Re: Simulated using Cholesky square root and the reformulated correction steps. Column labeled Sy+Re: Simulated using symmetric matrix square root and the reformulated correction steps.

Note that for the simulation results shown in Table 5.14 the constraints used are $\chi^x_{k-1} \geq 0$ (CC_1) in all the algorithms. For the simulations using the reformulated correction steps, the constraints $\chi^x_k \geq 0 (CC_7)$ are also

[13]Notation: UKF D.Simon means the algorithm in Table(5.4). UKF Non-aug means the algorithm in Table(5.5). UKF Q-aug means the algorithm in Table(5.6). UKF Fully-aug means the algorithm in Table(5.7).

Table 5.14: The table shows the average square root error based on 100 Monte Carlo simulations of the '2 state CSTR' case using constrained versions of the UKF algorithms.

	Ch	Sy	Ch+Re	Sy+Re
UKF D.Simon	$[0.37\ 0.43]^T$	$[0.36\ 0.41]^T$	$[0.32\ 0.42]^T$	$[0.32\ 0.42]^T$
UKF Non-aug	$[0.42\ 0.48]^T$	$[0.40\ 0.46]^T$	$[0.33\ 0.45]^T$	$[0.33\ 0.45]^T$
UKF Q-aug	$[0.49\ 0.55]^T$	$[0.45\ 0.51]^T$	$[0.41\ 0.51]^T$	$[0.40\ 0.49]^T$
UKF Fully-aug	$[0.52\ 0.58]^T$	$[0.46\ 0.52]^T$	$[0.44\ 0.53]^T$	$[0.42\ 0.52]^T$

used. The notation in the table is as follows: Column labeled QP Ch: Simulated using Cholesky square root and the QP version of the correction steps. Column labeled QP Sy: Simulated using Symmetric square root and the QP version of the correction steps. Column labeled NLP Ch: Simulated using Cholesky square root and the NLP version of the correction steps. Column labeled NLP Sy: Simulated using Symmetric matrix square root and the NLP version of the correction steps.

Table 5.15: The table shows the average square root estimation error based on 100 Monte Carlo simulations of the '2 state CSTR' case using constrained versions of the UKF algorithms.

	QP Ch	QP Sy	NLP Ch	NLP Sy
UKF D.Simon	$[0.34\ 0.40]^T$	$[0.34\ 0.41]^T$	$[0.34\ 0.41]^T$	$[0.34\ 0.41]^T$
UKF Non-aug	$[0.35\ 0.41]^T$	$[0.35\ 0.42]^T$	$[0.35\ 0.41]^T$	$[0.35\ 0.41]^T$
UKF Q-aug	$[0.41\ 0.48]^T$	$[0.40\ 0.46]^T$	$[0.42\ 0.48]^T$	$[0.41\ 0.48]^T$
UKF Fully-aug	$[0.45\ 0.52]^T$	$[0.43\ 0.50]^T$	$[0.45\ 0.52]^T$	$[0.43\ 0.50]^T$

Note also that for the simulation results shown in Table 5.13 the constraints used are $\chi^x_{k-1} \geq 0$ (CC_1) in all the algorithms. For the QP UKF's the cost function used is as by (5.33). For the NLP UKF's the cost function used is as by (5.32).

Table 5.11-5.13 indicates that adding complexity to the original UKF algorithms have a computational cost, but as Table 5.14-5.15 indicates, with a potential increase in estimation accuracy, even though the most complex not necessarily gives the most accurate estimation. Given a system, which algorithm to select is a trial and error process.

Chapter 6

Noise modeling concepts in nonlinear state estimation

This chapter is based on (Kolås, Foss and Schei 2008) as submitted[1] *to Journal of Process Control.*

6.1 Abstract

It is quite common to assume that uncertainty enters through additive white noise sources when using recursive state estimation algorithms. Also unknown and time-varying parameters are often modeled similarly by augmenting the states with a parameter vector. Further, it is common to reflect initial model uncertainty through the choice of the initial covariance matrices for the states and parameters.

In this paper we study noise modeling based on a hypothesis that it is important to model noise correctly. In practice this implies a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty. Alternative concepts of modeling the noise are investigated, and it is shown that modeling noise by introducing it in the system auxiliary variables and control inputs may have a positive impact on estimation performance.

Keywords: Nonlinear state estimation, Constrained nonlinear systems, Noise modeling, Joint UKF.

[1] Resubmitted version based on feedback from reviewers.

6.2 Introduction

Recursive state estimation algorithms usually assumes that uncertainty enters through additive white noise sources. Further, unknown and time-varying parameters are often modeled similarly by augmenting the states with a parameter vector. Finally, initial model uncertainty is reflected through the choice of the initial covariance matrices for the states and parameters.

Studying noise modeling based on a hypothesis that it is important to model noise correctly, implies in practice a critical view on the dominating 'additive noise paradigm' as a means to model uncertainty.

The 'additive noise paradigm' dominates textbooks and papers on recursive state estimation, i.e. Kalman filter type algorithms like the EKF and UKF (see e.g. Simon (2006), Søderstrøm and Stoica (1989), Gelb (1974)) even though uncertainty may enter a system in many different ways. The additive noise model structure is obviously reasonable in many applications. In others, however, this is not the case. One example are processes where control input uncertainty dominates, and where this noise depends on the value of the control input itself. It may for instance increase proportionally with the control input.

A fruitful way to view noise modeling is to view this as a direct extension of the process of developing a model. We assume dynamic models which are developed using physical insight and process data, i.e. physics-based models. Having established and possibly validated such a model, it is at least in principle possible to quantify uncertainty. This may include uncertainty in initial conditions, and in certain time-varying states and parameters, control inputs and measurements. Further, it may be possible to describe how noise enters the system, i.e. to structurally model how uncertainty affects the model.

It should be added that noise modeling has attracted more attention in the system identification literature (see e.g. Ljung (1987)). One example is how white noise may enter an input-output model in different ways.

In this paper alternative noise modeling concepts are investigated, and it is shown that introducing noise in auxiliary variables and control inputs may have a positive impact on estimation performance.

The paper is organized as follows. First noise modeling concepts are presented. Then follows a simulation study based on available literature examples and to an industrial process. Finally, some discussion and conclusions ends the paper.

6.3 System description

In this work we address the general continuos nonlinear system given by the state space formulation

$$\begin{aligned} \dot{x}(t) &= f_c(x(t), \theta(t), u(t), v(t)) \\ & \quad x(0) \text{ - given} \end{aligned} \tag{6.1}$$

$$y(t) = h(x(t), w(t)) \tag{6.2}$$

where $x(t)$ denotes the system states, $\theta(t)$ denotes the system parameters, $y(t)$ the output model (or measurements), $u(t)$ control inputs, $v(t)$ system state noise input (stationary or time-varying stochastic), $w(t)$ output (measurement) noise inputs (stationary or time-varying stochastic), and $f_c(\cdot)$ and $h(\cdot)$ are nonlinear Lipschitz continuos functions. All variables are vectors with appropriate dimensions.

Further we address the general discrete time nonlinear system with constant sampling time (denoted by k) given by the state space formulation

$$\begin{aligned} x_k &= f(x_{k-1}, \theta_{k-1}, u_{k-1}, v_{k-1}) \\ & \quad x_0 \text{ - given} \end{aligned} \tag{6.3}$$

$$y_{k-1} = h(x_{k-1}, w_{k-1}) \tag{6.4}$$

where x_k denotes the system states, θ_k denotes the system parameters, y_k the output model (or measurements), u_k control inputs, v_k system state noise input (stationary or time-varying stochastic), w_k output (measurement) noise inputs (stationary or time-varying stochastic), $f(\cdot)$[2] and $h(\cdot)$ are nonlinear functions.

All variables are vectors with appropriate dimensions. The model used in the estimators and estimated values are denoted with a $\hat{}$, typically $\hat{x}_k$, $\hat{y}_k$ etc.

We have chosen to present the analytical work using the continuos formulation, while all the simulations and estimations are done in discrete time. We assume that the analytical results valid for the continuos time case also are valid for the discrete time case.

[2]Note that the nonlinear function $f(\cdot)$ is typically a discretized representation of a nonlinear continous model $f_c(\cdot)$. The discretizition method used is typically Runge-Kutta 4^{th} order.

6.3.1 State estimation

In 1960 Rudolph Kalman presented a recursive state estimation method, which has become known as the Kalman filter (Kalman (1960)). The important contribution was the fact that a recursive algorithm could be used to accurately compute the first and second order moments (mean and covariance) of a linear system corrupted by Gaussian white noise on its system and output models. Kalman's simple assumptions were that the system random variables could be consistently estimated by sequentially updating the first and second order moments, and that the estimator was on the linear form

$$\hat{x}_k = \hat{x}_k^- + K_k \cdot e_k \tag{6.5}$$

The Kalman filter consists of two parts; a forward prediction part and a correction part. The prediction part computes a predicted (a prior) estimate of the first and second order moments at time k given information up until $k-1$ denoted by $\hat{x}_k^-$ and $P_{x_k}^-$. The correction part computes the corrected (posterior) estimates $\hat{x}_k$ and P_{x_k} using available data at time k. In (6.5) the predicted estimate ($\hat{x}_k^-$) is updated by a linear gain (K_k) times an error (e_k). The error part is the deviation between an output value(s) (y_k) and the estimate of the said output(s) ($\hat{y}_k$), and hence (6.5) can be formulated as

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - \hat{y}_k) \tag{6.6}$$

where the linear gain (K_k), or the so-called Kalman gain, is given by

$$K_k = P_{x_k y_k} P_{y_k y_k}^{-1} \tag{6.7}$$

where $P_{x_k y_k}$ and $P_{y_k y_k}$ is the cross covariance and output covariance, respectively.

The Extended Kalman filter (EKF), which was originally proposed by Stanley Schmidth in 1967 (Bellatoni and Dodge (1967)) in order to apply the Kalman filter to nonlinear spacecraft navigation problems, is probably the most used method in applied nonlinear state estimation. However, several authors have experienced shortcomings applying EKF to systems with severe nonlinearity and/or constraints (see for example Julier and Uhlmann (1994), Julier et al. (1995), Schei (1997), Nørgaard et al. (2000), Rao (2000), Hasseltine and Rawlings (2003), Bizup and Brown (2003), Chen et al. (2006), Vachhani et al. (2006) and Kandepu et al. (2008) to mention some). The shortcomings are related to difficulties in determining the Jacobians, errors introduced by linearization and/or to deal with systems

with multimodal or asymmetric probability density functions (pdf). Also, if handling of constraints are unavoidable, the EKF has some limitations in propagating the constraints both through the states and covariance calculations.

The Unscented Kalman[3] filter, proposed by Julier and Uhlmann (1994)[4], apply almost the same computational structure as the EKF, but is based on the intuition that "*it is easier to approximate a probability distribution than a nonlinear function*" (Julier et al. (2000), Julier (2002)). By using the Unscented transform to compute the mean and covariance, the Unscented Kalman filter avoids the need to use Jacobians in the algorithm. The probability distribution is approximated by a set of deterministic points which captures the mean and covariance of the distribution. These points, called sigma points, are then processed through the nonlinear model of the system, producing a set of propagated sigma points. By choosing appropriate weights, the weighted average and the weighted outer product of the transformed points gives the mean (for example $\hat{x}_k^-$) and covariance (for example $P_{x_k}^-$) of the transformed distribution. I.e. assume there exist a sigma point set χ_{k-1}^x with $2n$ sigma points, a set of appropriate weights W_i^x and W_i^c, and a nonlinear function (or transformation)

$$x_k = f(x_{k-1}) \tag{6.8}$$

Passing the vector of sigma points $\chi_{k-1,i}^x$ through the nonlinear function results in the propagated sigma points

$$\chi_{k,i}^x = f(\chi_{k-1,i}^x) \tag{6.9}$$

and the approximated mean $\hat{x}_k$ and the covariance P_{x_k} could be calculated as

$$\hat{x}_k = \sum_{i=1}^{2n} W_i^x \cdot \chi_{k,i}^x \tag{6.10}$$

$$P_{x_k} = \sum_{i=1}^{2n} W_i^c (\chi_{k,i}^x - \hat{x}_k)(\chi_{k,i}^x - \hat{x}_k)^T{}_k \tag{6.11}$$

[3] In the original work of Julier and Uhlmann (1994) the Unscented Kalman filter was named 'The New Filter'.

[4] We refer to this document as the first work of UKF, since it was submitted to IEEE Transactions on Automatic Control in 1994, although not accepted before in 2000.

In this paper the analysis is done using the continuos time formulation, and we assume these results also are valid for the discrete time case. Based on this assumption, and for the case of simplicity regarding the theoretical results presented, we investigate the covariance equation found in the continuos EKF (Gelb (1974))

$$\dot{P} = FP + PF^T + HQH^T - KRK^T \tag{6.12}$$

where all the elements in (6.12) may be time varying. It is common practice to look at the covariance (P) as if it reflects the confidence in the system states $\hat{x}$, and that quantitatively a relative high value means quite uncertain and that a relative small value means certain state estimates (Søderstrøm and Stoica, 1989, p. 326). In (6.12) F and H denote the Jacobians found by $F = \frac{\partial}{\partial x} f_c(x, \theta, u)|_{x=\hat{x},\theta=\hat{\theta},u,v}$, $H = \frac{\partial}{\partial v} f_c(x, \theta, u, v)|_{v,x,\theta,u}$. Especially the term HQH^T in (6.12) will have our attention in this work, since this is the term that determines how the uncertainty due to process noise is injected into (6.12) (Gelb, 1974, pp. 122). Further, we have the assumed process noise covariance matrix $E\left\{v(t)v^T(\tau)\right\} = Q(t)\delta(t-\tau)$ and the assumed measurement noise covariance matrix $E\left\{w(t)w^T(\tau)\right\} = R(t)\delta(t-\tau)$, where $\delta(t-\tau)$ is the Dirac function. Further we have the relationship between the continuos time spectral density Q and the discrete time covariance Q_k by

$$Q = Q_k/\Delta t \tag{6.13}$$

and R and the discrete measurement covariance matrix R_k by

$$R = R_k \Delta t \tag{6.14}$$

where Δt denotes the discrete time step (Simon, 2006, pp. 233).

6.4 Noise modeling concepts

As already mentioned in the *Introduction*, the literature concerning modeling and estimation of stochastic systems usually model the uncertainty or noise as additive to the system equations and/or output equations.

In the sections to follow we investigate the effect of different system noise modeling methods in the scope of the Kalman filter approach to estimation.

Describing the ideas covered, are in our opinion best done by introducing them by examples on the system introduced below, because we think it makes the ideas clearer than by using a general formulation.

6.4.1 Method 1 - Additive noise

Consider the gas-phase, reversible reaction Hasseltine and Rawlings (2003)

$$2A \overset{k_r}{\rightarrow} B \tag{6.15}$$

with stoichiometric matrix

$$s = \begin{bmatrix} s_1 & s_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \end{bmatrix} \tag{6.16}$$

and reaction rate

$$r = k_r P_A^2 \tag{6.17}$$

The state and output vectors are defined as

$$x = \begin{bmatrix} P_A \\ P_B \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad y = \begin{bmatrix} 1 & 1 \end{bmatrix} x \tag{6.18}$$

where P_A and P_B are the partial pressures. It is assumed that the ideal gas law holds (high temperature, low pressure), and that the reaction occurs in a well-mixed, isothermal batch reactor. Further we assume the reaction parameter k_r varies and is modelled as colored noise, and is estimated. The state and output vectors for the estimator are defined as

$$\hat{x} = \begin{bmatrix} \hat{P}_A \\ \hat{P}_B \\ \hat{k}_r \end{bmatrix} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix}, \qquad \hat{y} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \hat{x} \tag{6.19}$$

By assuming that the system (6.15)-(6.18) is considered and extended with control inputs (u) and by assuming additive noise on the system, we get the model for the process as

$$\dot{x_1} = -2r + u_1 + v_1 \tag{6.20}$$

$$\dot{x_2} = r + u_2 + v_2 \tag{6.21}$$

Described the traditional way by assuming additive noise on the system, we get the model for the estimator as

$$\dot{\hat{x}}_1 = -2\hat{r} + u_1 + v_1 \tag{6.22}$$

$$\dot{\hat{x}}_2 = \hat{r} + u_2 + v_2 \tag{6.23}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{6.24}$$

where the state vector $\hat{x}$ is given by (6.19) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r \hat{x}_1^2 = \hat{x}_3 \hat{x}_1^2 \tag{6.25}$$

The noise vector is given by

$$v = \begin{bmatrix} v_1 & v_2 & v_{k_r} \end{bmatrix}^T \tag{6.26}$$

where $v_1 \sim N(0, \bar{v}_1^2)$, $v_2 \sim N(0, \bar{v}_2^2)$ and $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$. The parameters are given by

$$\hat{\theta} = \hat{k}_r \tag{6.27}$$

Assume that the system (6.15)-(6.18) is considered as a semi-batch system in that the species B is removed and that the species A is refilled when a certain level of A is reached resulting in the control input described by (6.28)[5]

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \begin{Bmatrix} 4 - x_1 \\ -x_2 \end{Bmatrix} & x_1 \leq 0.2 \\ \\ \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} & otherwise \end{bmatrix} \tag{6.28}$$

The Jacobian H is given by the identity matrix I, and hence HQH^T in (6.12) becomes

$$HQH^T = \begin{bmatrix} \bar{v}_1^2 & 0 & 0 \\ 0 & \bar{v}_2^2 & 0 \\ 0 & 0 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{6.29}$$

As mentioned above, it is common practice to look at the covariance (P) as if it reflects the confidence in the system states $\hat{x}$, and that quantitatively a relative high value means quite uncertain and that a relative

[5] Note that the in the actual implementation, using a discrete formulation, the control input is scaled such that the refilling/removing is done in one sample.

small value means certain state estimates (Søderstrøm and Stoica, 1989, p. 326). A challenge using the proposed noise formulation might be that if the covariance has settled to a relative low value (i.e. certain state estimates) when the control inputs is applied, representing removing B and refilling A, this may introduce large errors in the estimates. Also worth noticing is that with this formulation there is no information about the control inputs in the covariance equation given by (6.12).

6.4.2 Method 2 - Noise in control inputs

Assume that the system (6.15)-(6.18) is considered a semi-batch system in that the species B is removed and that the species A is refilled when a certain low level of A (represented by the partial pressure P_A) is reached (as described by (6.28)). Consider also that there is some uncertainty related to the control inputs, and that the uncertainty in the inputs can be expressed as a relative uncertainty. The model for the estimator becomes

$$\dot{\hat{x}}_1 = -2\hat{r} + u_1(1 + v_{u_1}) \tag{6.30}$$

$$\dot{\hat{x}}_2 = \hat{r} + u_2(1 + v_{u_2}) \tag{6.31}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{6.32}$$

where the state vector $\hat{x}$ is given by (6.19) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r \hat{x}_1^2 = \hat{x}_3 \hat{x}_1^2 \tag{6.33}$$

The noise vector is given by

$$v = \begin{bmatrix} v_{u_1} & v_{u_2} & v_{k_r} \end{bmatrix}^T \tag{6.34}$$

where $v_{u_1} \sim N(0, \bar{v}_{u_1}^2)$, $v_2 \sim N(0, \bar{v}_{u_2}^2)$ and $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$. The parameters are given by (6.27). The Jacobian H becomes

$$H = \begin{bmatrix} u_1 & 0 & 0 \\ 0 & u_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.35}$$

Assuming Q is described by

$$Q = \begin{bmatrix} \bar{v}_{u_1}^2 & 0 & 0 \\ 0 & \bar{v}_{u_2}^2 & 0 \\ 0 & 0 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{6.36}$$

the term HQH^T in (6.12) becomes

$$HQH^T = \begin{bmatrix} u_1^2\bar{v}_{u_1}^2 & 0 & 0 \\ 0 & u_2^2\bar{v}_{u_2}^2 & 0 \\ 0 & 0 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{6.37}$$

Consider again that the covariance (P) reflects the uncertainty of the system states $\hat{x}$. In the case when the control inputs are applied, representing removing B and refilling A, large errors in the estimates may be introduced. This is reflected in the proposed formulation by the injection of the input uncertainty into the covariance function.

6.4.3 Method 3 - Noise in auxiliary variables

Consider the system (6.15)-(6.18) and that process knowledge suggest that the noise enters the system through the reaction rate r, instead of directly on the system states. That is

$$\dot{x}_1 = -2r \tag{6.38}$$

$$\dot{x}_2 = r \tag{6.39}$$

where the state vector x is given by (6.18) and the reaction rate r is given by

$$r = k_r x_1^2 + v_r \tag{6.40}$$

The model for the estimator becomes

$$\dot{\hat{x}}_1 = -2\hat{r} \tag{6.41}$$

$$\dot{\hat{x}}_2 = \hat{r} \tag{6.42}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{6.43}$$

where the state vector $\hat{x}$ is given by (6.19) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r\hat{x}_1^2 + v_r = \hat{x}_3\hat{x}_1^2 + v_r \tag{6.44}$$

The noise vector is given by

$$v = \begin{bmatrix} v_r & v_{k_r} \end{bmatrix}^T \tag{6.45}$$

where $v_r \sim N(0, \bar{v}_r^2)$ and $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$. The parameters are given by (6.27). The system becomes

$$\dot{\hat{x}}_1 = -2\hat{k}_r\hat{x}_1^2 - 2v_r \tag{6.46}$$

$$\dot{\hat{x}}_2 = \hat{k}_r\hat{x}_1^2 + v_r \tag{6.47}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{6.48}$$

and the Jacobian H

$$H = \begin{bmatrix} -2 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.49}$$

Assuming Q is described by

$$Q = \begin{bmatrix} \bar{v}_r^2 & 0 \\ 0 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{6.50}$$

the term HQH^T in (6.12) becomes

$$HQH^T = \begin{bmatrix} 4\bar{v}_r^2 & -2\bar{v}_r^2 & 0 \\ -2\bar{v}_r^2 & \bar{v}_r^2 & 0 \\ 0 & 0 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{6.51}$$

That is, by applying noise on the auxiliary variable $\hat{r}$, correlation is naturally introduced in the covariance calculation (P). This could also be seen as an alternative to off-diagonal tuning on Q in (6.29), in that the correlation enters the system naturally and correctly scaled in the off-diagonals.

6.5 Simulation studies

In the following sections results with some of the noise methods described above are presented. Note that modeling error may be introduced via θ_k and to some extent through the noise inputs v_k and w_k. Further, the noise inputs are generated using the randn() function in Matlab. Note also that in all the simulations a discrete representation based on Runge-Kutta 4^{th} order integration scheme is used.

The first case is from Hasseltine and Rawlings (2003), where it was used to indicate the convergence performance of the Moving Horizon Estimation (MHE) approach and at the same time show the limitations using the EKF. The key challenge here is that it has a multi modal probability density

function, where some solutions are physically valid, and some not. Given a bad initial guess, an algorithm based on the Kalman filter approach may or may not converge to the true state.

The second case shows results from simulations done on an industrial simulator of the Hall-Heroult process.

The third case is a reactor case, inspired by the problem described by Rawlings et al. (1989), investigating one of Kodak's semi-batch specialities chemicals for photographic film production. The case appears in Rao (2000), and is used to compare the performance of the EKF and MHE.

All simulations were done in Matlab Release 14. The estimator used in this work is the fully augmented UKF with reformulation of the correction steps and the use of constraints as presented in Kolås et al. (2008a), hereafter denoted CUKF. Further, when parameter estimation is required, the CUKF used is an extended version allowing parameter estimation (Joint CUKF[6]). Note that the UKF requires a discrete representation of the system.

To cope with singular/negative definite covariance matrix with respect to matrix square root calculations, we have introduced the term δ_i such that

$$\begin{aligned} P_{x_k} &= P^-_{x_k} - \delta_k K P^-_{y_k y_k} K^T \\ \Omega_i &= [0, 0.1, 0.2, 0.3, \cdots, 1] \\ \delta_k &= \max(\Omega_i)|_{P_{x_k} > 0} \end{aligned} \tag{6.52}$$

If $P^-_{x_k} > 0$ then so is P_{x_k}. Note also that scaling is not an issue in any of the following simulations.

[6] The augmented CUKF/UKF is in relevant litterature referred to as augmentation of the noise matrices in the sigma points, hence the use of Joint to state that parameters are augmented into the state vector.

6.5.1 Case '2 state CSTR'

Case description

In the following, the systems as described by (6.15)-(6.28) are investigated. Further, the parameters for these systems, if not otherwise noted, is

$$\begin{gathered}
\Delta t = t_{k+1} - t_k = 0.1 \\
P_0 = \begin{bmatrix} 6^2 & 0 & 0 \\ 0 & 6^2 & 0 \\ 0 & 0 & 0.015^2 \end{bmatrix} \\
k_r = 0.16 \\
x_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \\
\hat{x}_0 = \begin{bmatrix} 0.1 \\ 4.5 \\ 0.9k_r \end{bmatrix}
\end{gathered} \tag{6.53}$$

Note that the initial guess ($\hat{x}_0$) for the estimator is poor.

First investigated is a case where the reaction parameter k_r is constant. It has a wrong initial value, however. Then follows an investigation of a case with step changes in the reaction parameter k_r. The step changes in k_r are invoked as

$$\begin{array}{cccccc} t & 0 & 46 & 200 & 400 \\ k_r & 0.16 & 0.12 & 0.17 & 0.16 \end{array} \tag{6.54}$$

The following constraints are applied to the CUKF sigma points

$$\begin{gathered} \textit{Lower bounds}: \ [0, 0, 0.1]^T \\ \textit{Upper bounds}: [\infty, \ \infty, \ , 0.18]^T \end{gathered} \tag{6.55}$$

Note that the actual CUKF applied is a Joint CUKF, since the state vector is augmented in order to perform parameter estimation of k_r. Note also that for all the simulations the simulator is based on a discrete representation of (6.20)-(6.21).

Simulation results with constant k_r

Method 1 The system equations and the noise is modeled as given in the description of Method 1, the estimator constraints as (6.55) and the estimator tuning as

$$\begin{aligned} Q_k &= diag\left[\begin{array}{ccc} \bar{v}_{k,1}^2 & \bar{v}_{k,2}^2 & \bar{v}_{k,k_r}^2 \end{array}\right] \\ &= diag\left[\begin{array}{ccc} \left(10^{-9}\right)^2 & \left(10^{-9}\right)^2 & \left(10^{-4}\right)^2 \end{array}\right] \\ R_k &= \bar{w}_k^2 = 0.002^2 \end{aligned} \tag{6.56}$$

In the case when the true process, i.e. the simulator, experience no noise, we get the results as shown in Figure 6.1.

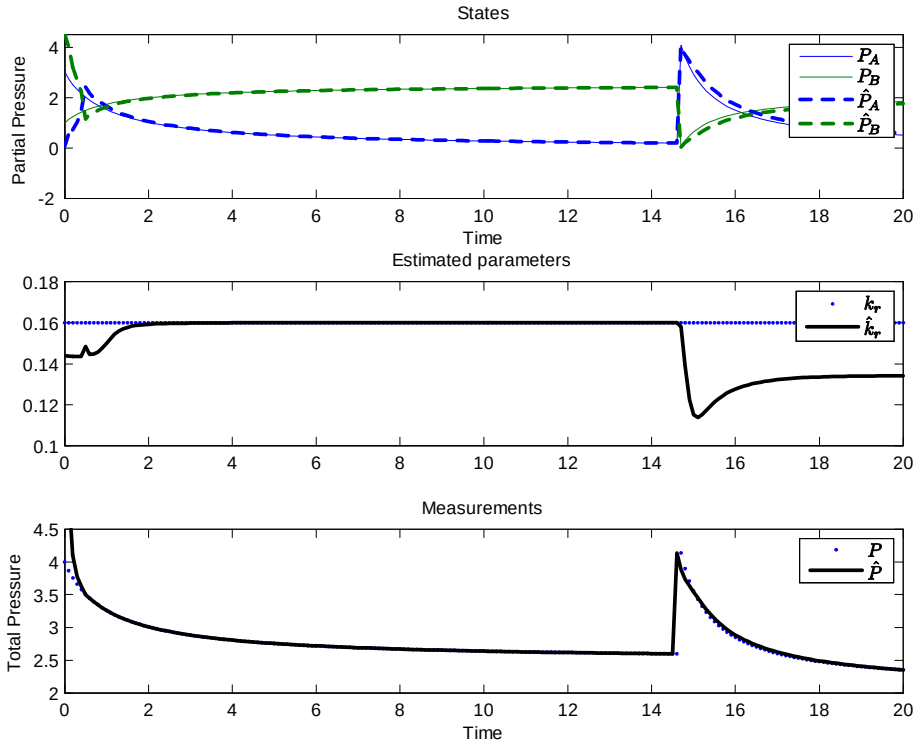

Figure 6.1: The figure shows the true and estimated states, the true and estimated reaction 'constant' k_r and the true and estimated output using Method 1. As the figure shows, the estimates of k_r is disturbed in the transient period caused by the control input excitation.

As Figure 6.1 shows, the bad initial guess is handled very well, and the state estimate and output (measurement) estimate is generally acceptable. However the parameter estimate during the transient period at approximately 'Time 15' is generally bad due to the control input excitation. In Figure 6.2 the time trace of the covariance is shown, and as expected the control input excitation is not reflected in the covariance.

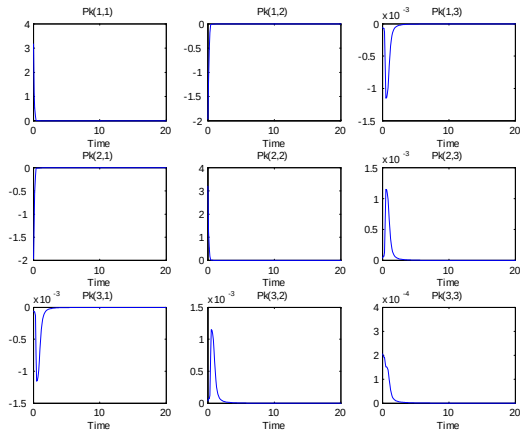

Figure 6.2: The time trace of the discrete covariance matrix P_{x_k}.

Method 1 and 2 combined By combining the concept described by Method 1 and 2, and by that introduce uncertainty also in the control inputs, the system equations for the estimator becomes

$$\dot{\hat{x}}_1 = -2\hat{r} + u_1(1 + v_{u_1}) + v_1 \tag{6.57}$$

$$\dot{\hat{x}}_2 = \hat{r} + u_2(1 + v_{u_2}) + v_2 \tag{6.58}$$

$$\dot{\hat{x}}_3 = v_{k_r} \tag{6.59}$$

where the state vector $\hat{x}$ is given by (6.19) and the reaction rate $\hat{r}$ is given by

$$\hat{r} = \hat{k}_r \hat{x}_1^2 = \hat{x}_3 \hat{x}_1^2 \tag{6.60}$$

The noise is given by $v_1 \sim N(0, \bar{v}_1^2)$, $v_2 \sim N(0, \bar{v}_2^2)$, $v_{k_r} \sim N(0, \bar{v}_{k_r}^2)$, $v_{u_1} \sim N(0, \bar{v}_{u_1}^2)$, $v_{u_2} \sim N(0, \bar{v}_{u_2}^2)$ and the parameters as by (6.27). Further, by assuming that Q is described by

$$Q = \begin{bmatrix} \bar{v}_1^2 & 0 & 0 & 0 & 0 \\ 0 & \bar{v}_2^2 & 0 & 0 & 0 \\ 0 & 0 & \bar{v}_{k_r}^2 & 0 & 0 \\ 0 & 0 & 0 & \bar{v}_{u_1}^2 & 0 \\ 0 & 0 & 0 & 0 & \bar{v}_{u_1}^2 \end{bmatrix} \tag{6.61}$$

the term HQH^T in (6.12) becomes

$$HQH^T = \begin{bmatrix} \bar{v}_1^2 + u_1^2\bar{v}_{u_1}^2 & 0 & 0 \\ 0 & \bar{v}_2^2 + u_2^2\bar{v}_{u_2}^2 & 0 \\ 0 & 0 & \bar{v}_{k_r}^2 \end{bmatrix} \tag{6.62}$$

and as seen by (6.62), the control inputs will be present in the covariance equation.

With the estimator constraints as by (6.55), and the estimator tuning as

$$\begin{aligned} Q_k &= diag\begin{bmatrix} \bar{v}_{k,1}^2 & \bar{v}_{k,2}^2 & \bar{v}_{k,k_r}^2 & \bar{v}_{k,u_1}^2 & \bar{v}_{k,u_2}^2 \end{bmatrix} \\ &= diag\begin{bmatrix} \left(10^{-9}\right)^2 & \left(10^{-9}\right)^2 & \left(10^{-4}\right)^2 & 1^2 & 1^2 \end{bmatrix} \\ R_k &= \bar{w}_k^2 = 0.002^2 \end{aligned} \tag{6.63}$$

the results for the case when the simulator experience no noise, is as shown in Figure 2.2.

As Figure 6.3 shows, the estimate of the parameter k_r does not suffer from the control input excitation. However the 'cost' is some loss off accuracy regarding the state estimates in the first samples after the control input excitation. In Figure 6.4 the time trace of the covariance is shown, and as expected the control input excitation at approximately 'Time 15' is reflected in the covariance elements ($P_{x_k}(1,1)$, $P_{x_k}(1,2)$, $P_{x_k}(2,1)$, $P_{x_k}(2,2)$).

That is, by applying the idea that the covariance should reflect the uncertainty in the states around the control input excitation, good results are achieved by combining the concept in Method 1 and 2.

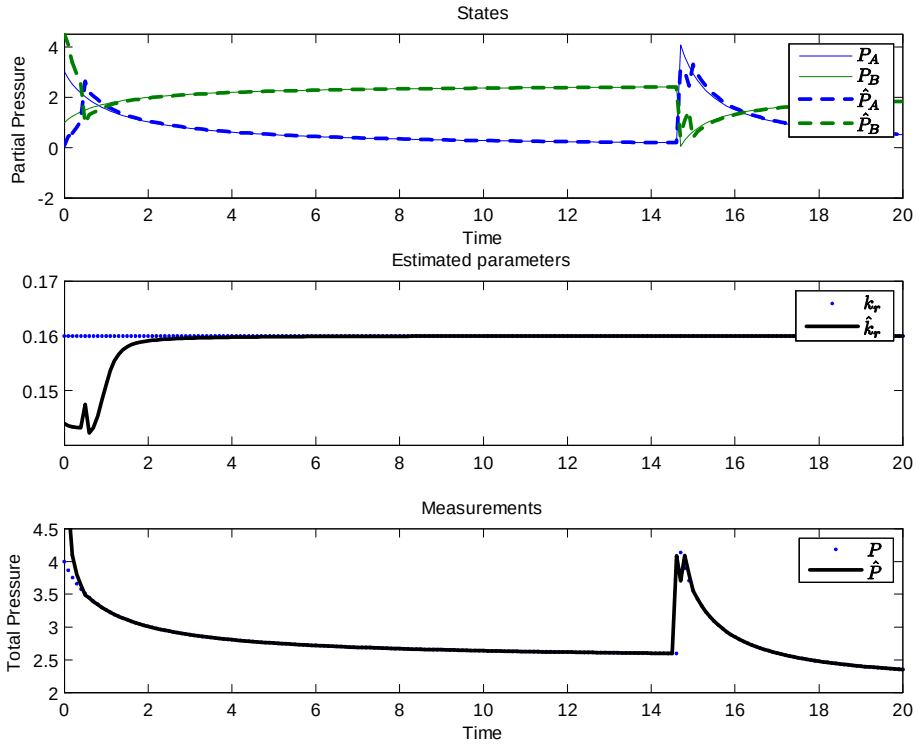

Figure 6.3: The figure shows the true and estimated states, the true and estimated reaction 'constant' k_r and the true and estimated output. As the figure shows the estimated k_r is not disturbed in the transient period caused by the control input excitation.

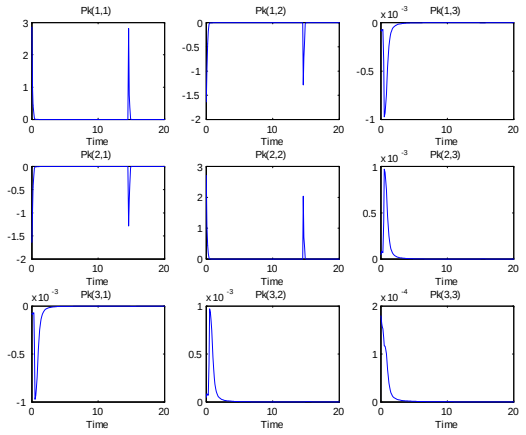

Figure 6.4: The time trace of the discrete covariance matrix P_{x_k}.

Simulation results with time-varying k_r

Method 1 Again we consider the system given by (6.22)-(6.28). That is, the traditional noise model (Method 1) and with the tuning of the estimator as in (6.55)-(6.56). The constraints regarding the sigma points χ^{x-}_{k-1} and χ^{x}_{k} is as in the case previously described. Further the parameter k_r changes according to (6.54). The result of the simulation is given in Figure 6.5.

As Figure 6.5 shows, the changes in the true reaction parameter k_r is to some extent reflected, but with a bias and some noise from the input excitation. However, Figure 6.5 reflects the case when the true process experiences no noise. By letting the true system experiences noise as

$$v \sim N\left(0^T, \begin{bmatrix} \bar{v}_1^2 & \bar{v}_2^2 & \bar{v}_{k_r}^2 & \bar{v}_{u_1}^2 & \bar{v}_{u_2}^2 \end{bmatrix}^T\right) = \tag{6.64}$$
$$N\left(0^T, \begin{bmatrix} 0.001^2 & 0.001^2 & 0.001^2 & 0.1^2 & 0.1^2 \end{bmatrix}^T \Delta t^{-1}\right)$$

$$w \sim N(0, \bar{w}^2) = N(0, 0.01^2 \Delta t) \tag{6.65}$$

and the estimator constraints as by (6.55) and estimator tuning as given by (6.56), the results are as shown in Figure 6.6.

As Figure 6.6 shows, the estimator performs bad given this tuning and noise sequence.

Method 1 and 2 combined By combining Method 1 and 2, i.e. consider the system with additive noise and assume relative noise in the control

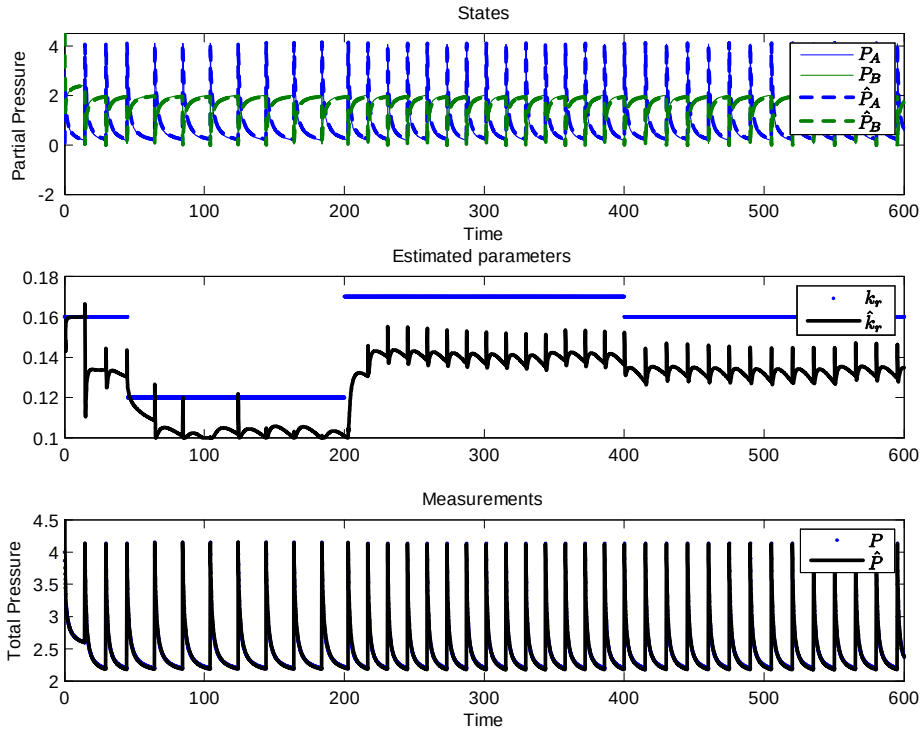

Figure 6.5: The figure shows the performance of the estimator using the noise formulation given in Method 1. The errors in the states are relative small, but the performance is bad with respect to how well the estimator suppresses the control input excitation and the large offset in the estimation of k_r.

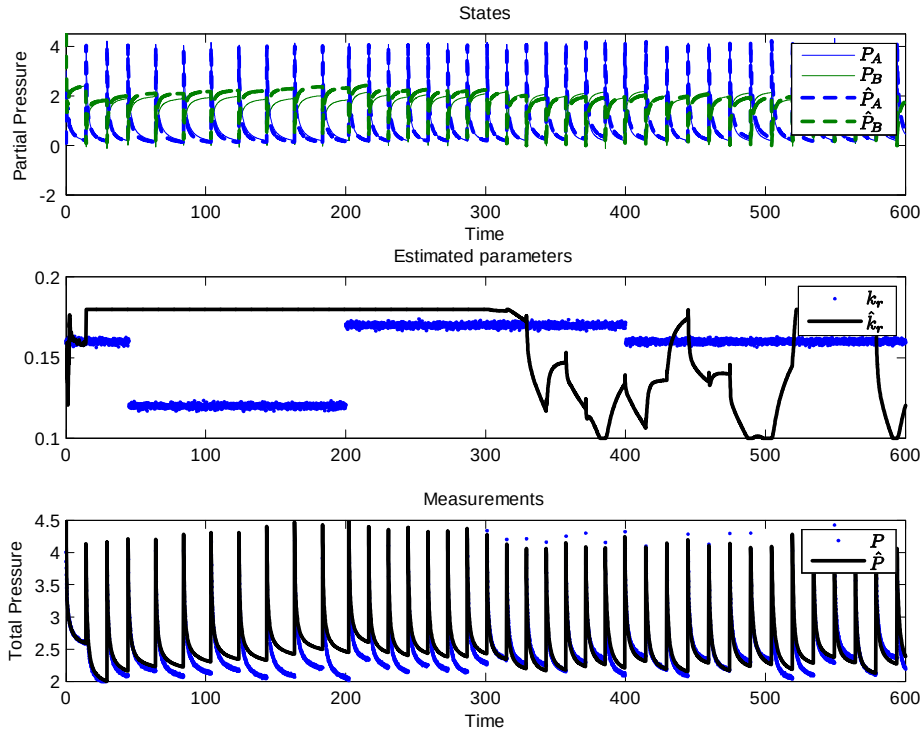

Figure 6.6: The figure shows the performance of the estimator using the noise formulation given in Method 1. The errors in the states are large, and the performance is bad with respect to the accuracy of the estimated k_r as well as poor suppression of the control input excitations.

inputs to let the covariance reflect the uncertainty in the state estimates when at the control input excitation. The estimator tuning is as given by (6.55) and (6.63). The results are shown in Figure 6.7.

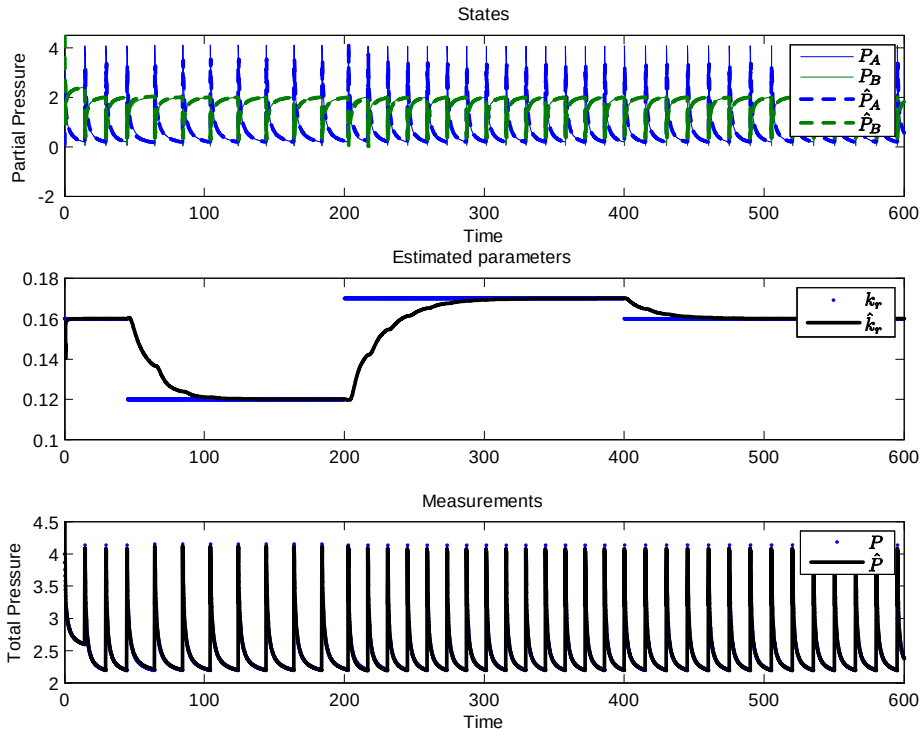

Figure 6.7: The figure shows the performance of the estimator using the noise formulation combining Method 1 and 2. The errors in the state estimates are small, and the performance good with respect to how well the estimator suppress the control input excitation and how well the estimation of k_r converge to the true state.

As Figure 6.7 shows the performance in the noise free case is acceptable. By letting the true system experience noise as given by (6.64)-(6.65), and the estimator tuning as by (6.55) and (6.63), we get the results as given by Figure 6.8.

As Figure 6.8 shows, given this particular noise sequence and particular noise levels, the estimator performs well with respect to the state estimates. The estimate of the parameter k_r is reasonably well, converging slowly towards the true state. Also, the control input excitations are no longer

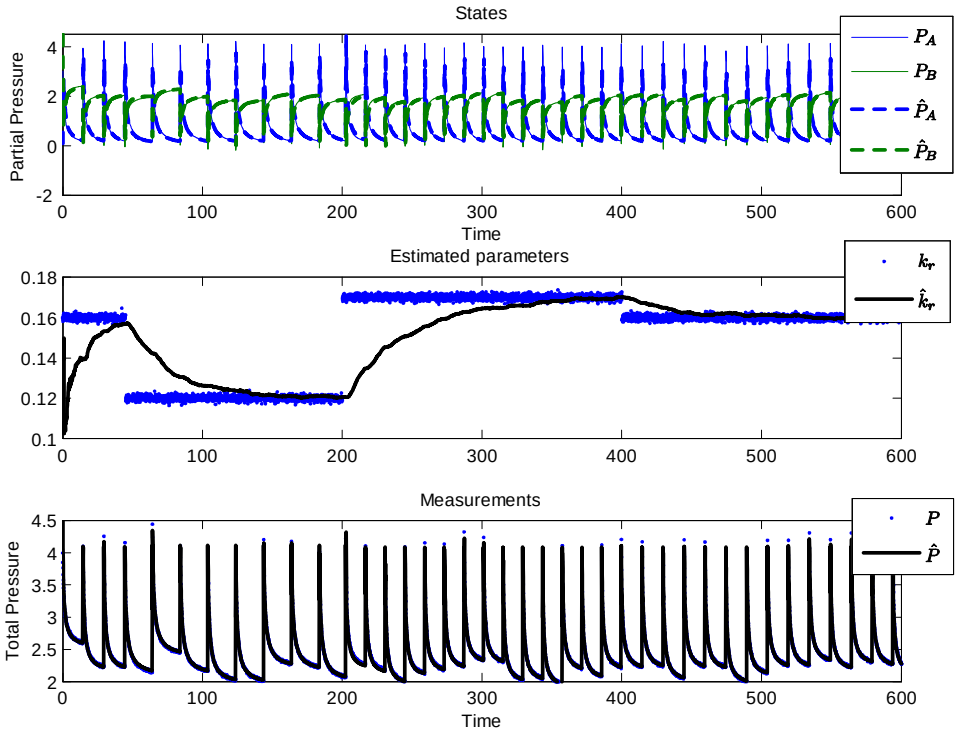

Figure 6.8: The figure shows the performance of the estimator using the noise formulation combining Method 1 and 2. The errors in the state estimates are relatively small, and the performance good with respect to how well the estimator suppress the control input excitation. Also, the estimator manages to estimate k_r reasonably well.

disturbing the estimate of k_r.

6.5.2 Case 'Hall-Heroult Process'

Case description

[7]The Hall-Heroult process is the dominating process for producing aluminum today (Grjotheim and Kvande (1993)). The fundamentals of the process are to dissolve Al_2O_3 in molten cryolite (also known as electrolyte or bath), and electrically reduce complex aluminum containing ions to pure aluminum. The overall electrochemical reaction in the electrolyte is

$$2Al_2O_3 + 3C \rightarrow 4Al + 3CO_2$$

where carbon is fed to the reaction as consumable anodes. By the use of various additives, in particular AlF_3, the operating temperature of the electrolyte can be lowered from 1010^oC to approximately 960^oC. Both decreased temperature and increased excess AlF_3 is believed to be beneficial for the amount of metal produced (current efficiency) and the energy consumption. As molten cryolite is very corrosive, the only component of an acceptable cost presently capable of coexisting with it over time is frozen cryolite. It is therefore necessary to maintain a layer of frozen cryolite (side ledge) to prevent the carbon walls from eroding. In order to maintain the side ledge there has to be a substantial heat loss through the side ledge and the carbon walls of the cell. The cell voltage applied is typically between 4.2 and 4.5V , and the electric current through the cell is typically 150 - 350kA. A sketch of a cell is shown in Figure 6.9. In a modern plant of today 100-300 cells are placed and connected in series. There are three control inputs to the process, anode beam adjustments (controlling energy input), addition of AlF_3 and addition of Al_2O_3, and three controlled variables, electrolyte temperature[8], concentration (or mass) of AlF_3 and concentration of Al_2O_3. A cell is regularly excited since liquid aluminium is tapped and some of the anode blocks are changed on a daily basis. This induces severe disturbances in the energy balance, and it implies that the operating conditions will vary significantly and hence provoke nonlinear cell effects. The process has strong internal couplings, for instance between the mass and energy balance through the side ledge. The coupled mass and energy balance combined with nonlinear process characteristics and few measure-

[7]The case description is mainly from Foss and Schei (2007)

[8]Also known as bath temperature.

ments, makes the Hall-Heroult process challenging to control (Foss and Schei (2007), Drengstig et al. (1998), Gran (1980)).

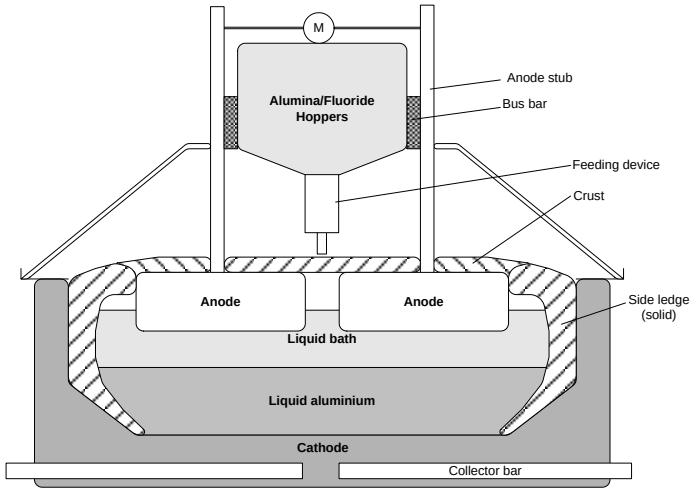

Figure 6.9: The figure illustrates a cell for producing liquid aluminium.

In this work we investigate, using an industrial simulator, the Hall-Heroult process in closed loop control by an advanced process control application as used by Hydro Aluminium AS. The model used is an adoption of the models found in Saksvikrønning et al. (1976) and Gran (1980). The estimators used in the simulations are the EKF (Gran (1980)) and the fully augmented UKF (van der Merwe (2004)).

As already mentioned, a particular challenge with the Hall-Heroult process is that the control inputs may experience severe disturbances. In the simulation a reduction of 5% in one of the control inputs is introduced in the simulator. The drop in the control input is not known to the control application. The purpose of the simulation is to study the noise disturbance models described by Method 1 and 2 on this process.

Simulation results

Figure 6.10 a) - c) shows the results of the simulation. The 5% drop in the control input is initiated at time stamp 1 and last until time stamp 6. After time stamp 6 the control input is considered back to normal. Note that the data is anonymized due to confidentiality issues.

Figure 6.10 indicates that the drop in the true control input at time step

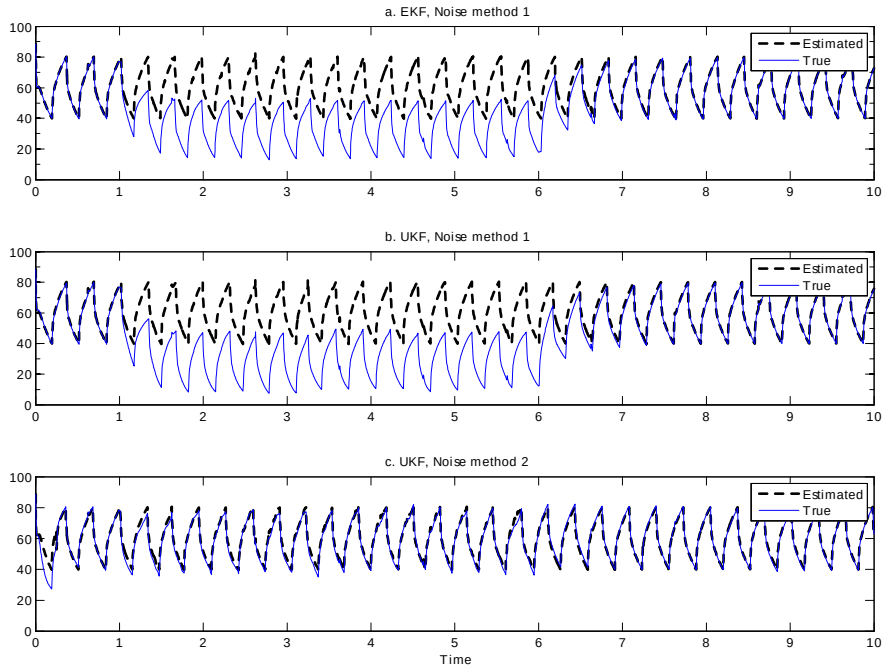

Figure 6.10: The figure indicates the effect of using the noise concepts as described by Method 1 (subplot a and b) and the combination of Method 1 and 2 (subplot c) on a simulated industrial process which experience input uncertainty.

1 is not 'discovered' by the estimator, either using EKF or UKF applying the noise modeling concept described by Method 1. However applying the noise modeling concept by combining Method 1 and 2 (i.e. input uncertainty), the UKF estimator 'discover' the drop, which is indicated in Figure 6.10 subplot c, by no drop in the **true** state (remember the system is in closed loop control). As indicated by Figure 6.10 subplot a and b, the drop in the control input is not discovered by the estimators using only the traditional additive noise formulation. Note that a simulation regarding how the EKF combined with noise combining Method 1 and 2 in this case was not performed, but is expected to behave similar to the UKF as shown in Figure 6.10, subplot c.

6.5.3 Case '4 State CSTR'

Case description

Consider a stirred tank reactor where the following liquid phase exothermic reaction occurs

$$A + 2B \rightarrow C. \tag{6.66}$$

This case is found in Rao (2000)[9], which is inspired by the problem described by Rawlings et al. (1989). The state estimation problem is to estimate precisely the concentration of A in the reactor. Because over addition of B leads to product degradation, precise concentration estimates of A as a function of time are necessary to complete the reaction without over-addition of B. Only temperature measurements corrupted with sensor noise are available. Further, the exact reaction kinetics are supposed unknown with the exception of the heat of reaction ΔH_r (Schuler and Schmidt (1992)). Under standard assumptions, such as negligible potential and kinetic energy effects, constant density, uniformly homogeneous mixture, and no phase transition, the reactor is simulated using the following model:Vachhani et al. (2005)

[9] Notet that the sampling time ($h = 0.1$ $[min]$) used in this work differ from the one ($h = 0.5$ $[min]$) used in Rao (2000), beacuse the simulator used in this work expeience instabilities if $h > 0.25$ $[min]$.

$$\dot{V} = F \tag{6.67}$$

$$\dot{A} = -k_0 e^{(-\frac{ER}{T})} AB^2 - \frac{F}{V}A \tag{6.68}$$

$$\dot{B} = -2k_0 e^{(-\frac{ER}{T})} AB^2 + \frac{F}{V}\left(C_{B_f} - B\right) \tag{6.69}$$

$$\dot{T} = -\frac{\Delta H_r}{\rho C_p} k_0 e^{(-\frac{ER}{T})} AB^2 + \frac{F}{V}\left(T_f - T\right) + \frac{UA}{\rho C_p V}\left(T_c - T\right). \tag{6.70}$$

The model parameters are listed in Table 6.1.

Table 6.1: The table explains the parameters used in the 4-state CSTR case.

Parameter.	Value	Units
k_0	$9 \cdot 10^{11}$	$\frac{L^2}{mol^2 \min}$
ER	6000	K
ρ	1000	$\frac{g}{L}$
C_p	0.239	$\frac{J}{g \cdot K}$
UA	$2 \cdot 10^5$	$\frac{J}{\min \cdot K}$
T_f	300	K
T_c	300	K
C_{B_f}	2.2	$\frac{mol}{L}$
ΔH_r	$-5 \cdot 10^4$	$\frac{J}{mol}$
F	-	$\frac{L}{\min}$
r	-	$\frac{mol}{\min \cdot L}$
Q_r	-	$\frac{K}{\min}$
$V(0)$	100	L
$A(0)$	0.5	$\frac{mol}{L}$
$B(0)$	0	$\frac{mol}{L}$
$T(0)$	300	K

The flowrate profile (F) used, is a rescaled version of the one used in the operation of the industrial reactor described by Rawlings et al. (1989). To account for imperfect cooling and modeling inaccuracies the cooling water temperature (T_c) is assumed fluctuating. The control inputs[10] $F(u_1)$ and the measured disturbance $T_c(u_2)$ is as given in Figure 6.11.

[10] Note that the control input u_1 is a scaled version of the one given in Rawlings et al. (1989). The 'measured' temperature of the coolingwater $T_c \sim N(\bar{T}_c, v_{T_c}^2) = N(300, 5)$ is generated based on the original sampling rate as u_1 and then scaled as in Rao (2000).

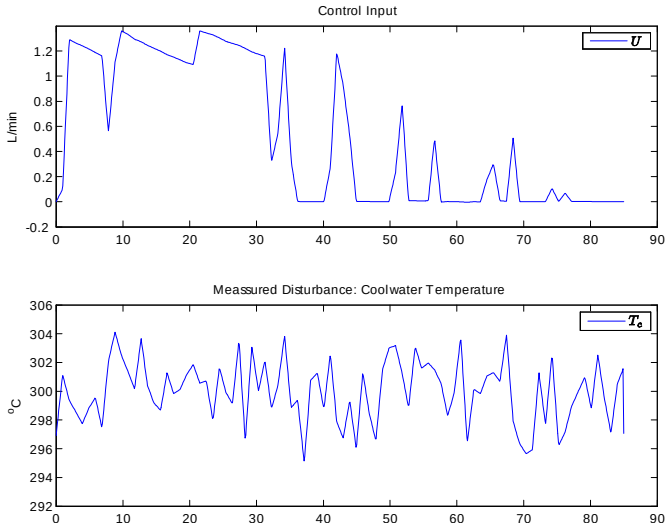

Figure 6.11: The control input u_1 and the measured cooling water temperature T_c.

By introducing the notation for the states x and control inputs u as

$$x = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}^T = \begin{bmatrix} V & A & B & T \end{bmatrix}^T \tag{6.71}$$

$$u = \begin{bmatrix} u_1 & u_2 \end{bmatrix}^T = \begin{bmatrix} F & T_c \end{bmatrix}^T, \tag{6.72}$$

the model can be formulated as

$$\dot{x}_1 = u_1 \tag{6.73}$$

$$\dot{x}_2 = -k_0 e^{(-\frac{ER}{x_4})} x_2 x_3^2 - \frac{x_2}{x_1} u_1 \tag{6.74}$$

$$\dot{x}_3 = -2k_0 e^{(-\frac{ER}{x_4})} x_2 x_3^2 + \frac{(C_{B_f} - x_3)}{x_1} u_1 \tag{6.75}$$

$$\dot{x}_4 = -\frac{\Delta H_r}{\rho C_p} k_0 e^{(-\frac{ER}{x_4})} x_2 x_3^2 + \frac{UA}{\rho C_p} \frac{(u_2 - x_4)}{x_1} + \frac{(T_f - x_4)}{x_1} u_1. \tag{6.76}$$

Defining the reaction rate r as

$$r = k_0 e^{(-\frac{ER}{x_4})} x_2 x_3^2 \tag{6.77}$$

gives the model

$$\dot{x}_1 = u_1 \tag{6.78}$$
$$\dot{x}_2 = -r - \frac{x_2}{x_1}u_1 \tag{6.79}$$
$$\dot{x}_3 = -2r + \frac{\left(C_{B_f} - x_3\right)}{x_1}u_1 \tag{6.80}$$
$$\dot{x}_4 = -\frac{\Delta H_r}{\rho C_p}r + \frac{UA}{\rho C_p}\frac{(u_2 - x_4)}{x_1} + \frac{(T_f - x_4)}{x_1}u_1. \tag{6.81}$$

Further it is assumed that the reaction rate r is unknown and we will investigate two different estimation approaches addressing this issue. Both approaches assumes that the reaction rate r could be modeled as colored noise. First we assume that only a simplified model of the reaction rate is available (as in Rao (2000)), and that the reaction rate r is derived from the estimated reaction energy Q_r. The estimator model takes the form

$$\dot{\hat{x}}_1 = u_1 \tag{6.82}$$
$$\dot{\hat{x}}_2 = -\hat{r} - \frac{\hat{x}_2}{\hat{x}_1}u_1 \tag{6.83}$$
$$\dot{\hat{x}}_3 = -2\hat{r} + \frac{\left(C_{B_f} - \hat{x}_3\right)}{\hat{x}_1}u_1 \tag{6.84}$$
$$\dot{\hat{x}}_4 = -\frac{\Delta H_r}{\rho C_p}\hat{r} + \frac{UA}{\rho C_p}\frac{(u_2 - \hat{x}_4)}{\hat{x}_1} + \frac{(T_f - \hat{x}_4)}{\hat{x}_1}u_1 \tag{6.85}$$
$$\dot{\hat{x}}_5 = v_{x_5} \tag{6.86}$$

where

$$\hat{x}_5 = \hat{Q}_r = -\frac{\Delta H_r}{\rho C_p}\hat{r} \tag{6.87}$$

and control inputs u as (6.72) and the states $\hat{x}$ as

$$\hat{x} = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & \hat{x}_4 & \hat{x}_5 \end{bmatrix}^T = \begin{bmatrix} \hat{V} & \hat{A} & \hat{B} & \hat{T} & \hat{Q}_r \end{bmatrix}^T \tag{6.88}$$

Next, we assume that knowledge on how the states enter the reaction rate (6.77) is available, but where the true value of the reaction rate constant

k_0 is uncertain and must be estimated. In this case $\hat{x}_5$ in (6.86) takes the form

$$\hat{x}_5 = \hat{k}_0 \tag{6.89}$$

and the reaction rate by

$$\hat{r} = \hat{k}_0 e^{(-\frac{ER}{\hat{x}_4})} \hat{x}_2 \hat{x}_3^2 \tag{6.90}$$

and the control inputs u as (6.72) and the states $\hat{x}$ as

$$\hat{x} = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & \hat{x}_4 & \hat{x}_5 \end{bmatrix}^T = \begin{bmatrix} \hat{V} & \hat{A} & \hat{B} & \hat{T} & \hat{k}_0 \end{bmatrix}^T \tag{6.91}$$

Note that for all the simulations to come, it is assumed that the true system experiences noise in x_1 and that the noise in the other states enters via noise in the reaction rate r, i.e.

$$r = k_0 e^{(-\frac{ER}{x_4})} x_2 x_3^2 + v_r \tag{6.92}$$

where

$$v_r \sim N(0, \bar{v}_r^2) = N\left(0, 10^{-6}/\Delta t\right) \tag{6.93}$$

$$v_{x_1} \sim N(0, \bar{v}_{x_1}^2) = N(0, 10^{-10}/\Delta t) \tag{6.94}$$

Further the true measurement noise is assumed to be

$$w \sim N(0, \bar{w}_T^2) = N\left(0, 0.1^2 \Delta t\right) \tag{6.95}$$

Method 1, simplified model

We investigate the behavior of the estimator when erroneously assuming the standard approach that all the states experience additive noise, and no noise in the auxiliary parameters. The state vector is as given by (6.88) and the estimator tuning as

$$\hat{x}_0 = \begin{bmatrix} 100 & 0.5 & 0 & 300 & 0 \end{bmatrix}^T \tag{6.96}$$

$$P_0 = diag\begin{bmatrix} 10^{-5} & 10^{-5} & 10^{-5} & 10^{-5} & 10^{-1} \end{bmatrix} \tag{6.97}$$

$$\begin{aligned} Q_k &= diag\begin{bmatrix} \bar{v}_{x_1}^2 & \bar{v}_{x_2}^2 & \bar{v}_{x_3}^2 & \bar{v}_{x_4}^2 & \bar{v}_{x_5}^2 \end{bmatrix} \\ &= diag\begin{bmatrix} \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(7 \cdot 10^{-2}\right)^2 & \left(10^{-5}\right)^2 & 1^2 \end{bmatrix} \end{aligned} \tag{6.98}$$

$$R_k = \bar{w}_k^2 = 0.075^2 \tag{6.99}$$

$$\begin{aligned} &\textit{Sigma points lower bounds} : [0,\ 0,\ 0,\ 0,\ 0]^T \\ &\textit{Sigma points upper bounds} : [\infty,\ \infty,\ 0.011,\ \infty,\ 3]^T \end{aligned} \tag{6.100}$$

Using the standard approach with additive noise on the states, the results are as given in Figure 6.12.

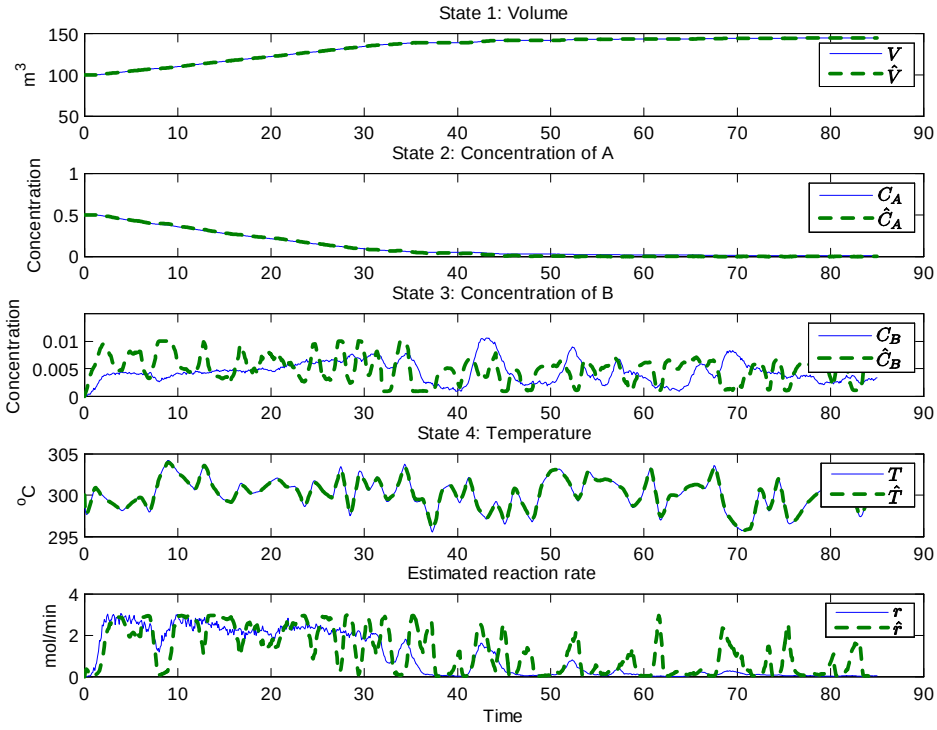

Figure 6.12: The results of the simulation assumed simplified model and additive noise.

As is seen by Figure 6.12, the estimation of the concentration of A is close to the true one, but the estimated reaction rate and the estimated concentration of B suffer from noise. In this case we have used v_{x_3} to tune the behavior of the estimated concentration of B.

Method 1 and 3 combined, simplified model

Applying 'process knowledge' we may assume that the reaction rate r is uncertain and experience noise as

$$\hat{Q}_r = -\frac{\Delta H_r}{\rho C_p}\left(\hat{r} + v_r\right) \tag{6.101}$$

$v_r \sim N(0, \bar{v}_r^2)$ and the reaction rate becomes

$$\hat{r} = -\hat{Q}_r \left(\frac{\Delta H_r}{\rho C_p}\right)^{-1} - v_r = -c_1 \hat{x}_5 - v_r \tag{6.102}$$

$$c_1 = \left(\frac{\Delta H_r}{\rho C_p}\right)^{-1} \tag{6.103}$$

The model for the estimator becomes

$$\dot{\hat{x}}_1 = u_1 + v_{x_1} \tag{6.104}$$

$$\dot{\hat{x}}_2 = -\hat{r} - \frac{\hat{x}_2}{\hat{x}_1} u_1 \tag{6.105}$$

$$\dot{\hat{x}}_3 = -2\hat{r} + \frac{\left(C_{B_f} - \hat{x}_3\right)}{\hat{x}_1} u_1 \tag{6.106}$$

$$\dot{\hat{x}}_4 = -c_1^{-1}\hat{r} + \frac{UA}{\rho C_p}\frac{(u_2 - \hat{x}_4)}{\hat{x}_1} + \frac{(T_f - \hat{x}_4)}{\hat{x}_1} u_1 \tag{6.107}$$

$$\dot{\hat{x}}_5 = v_{x_5} \tag{6.108}$$

The noise vector v is defined as

$$v = \begin{bmatrix} v_{x_1} & v_r & v_{x_5} \end{bmatrix}^T \tag{6.109}$$

where $v_{x_1} \sim N(0, \bar{v}_{x_1}^2)$, $v_r \sim N(0, \bar{v}_r^2)$ and $v_{x_5} \sim N(0, \bar{v}_{x_5}^2)$. By this H becomes

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & c_1^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.110}$$

and HQH^T

$$HQH^T = \begin{bmatrix} \bar{v}_{x_1}^2 & 0 & 0 & 0 & 0 \\ 0 & \bar{v}_r^2 & 2\bar{v}_r^2 & c_1^{-1}\bar{v}_r^2 & 0 \\ 0 & 2\bar{v}_r^2 & 4\bar{v}_r^2 & 2c_1^{-1}\bar{v}_r^2 & 0 \\ 0 & c_1^{-1}\bar{v}_r^2 & 2c_1^{-1}\bar{v}_r^2 & c_1^{-1}\bar{v}_r^2 & 0 \\ 0 & 0 & 0 & 0 & \bar{v}_{x_5}^2 \end{bmatrix} \tag{6.111}$$

when

$$Q = diag \begin{bmatrix} \bar{v}_{x_1}^2 & \bar{v}_r^2 & \bar{v}_{x_5}^2 \end{bmatrix} \tag{6.112}$$

For the estimator the state vector is as given by (6.88) and the other initial values are

$$\hat{x}_0 = \begin{bmatrix} 100 & 0.5 & 0 & 300 & 0 \end{bmatrix}^T \tag{6.113}$$

$$\begin{gathered} Sigma\ points\ lower\ bounds : [0,\ 0,\ 0,\ 0,\ 0]^T \\ Sigma\ points\ upper\ bounds : [\infty,\ \infty,\ 0.011,\ \infty,\ 3]^T \end{gathered} \tag{6.114}$$

$$P_0 = diag \begin{bmatrix} 10^{-5} & 10^{-5} & 10^{-5} & 10^{-5} & 10^{-1} \end{bmatrix} \tag{6.115}$$

$$\begin{aligned} Q_k &= diag \begin{bmatrix} \bar{v}_{k,x_1}^2 & \bar{v}_{k,r}^2 & \bar{v}_{k,x_5}^2 \end{bmatrix} \\ &= diag \begin{bmatrix} \left(10^{-5}\right)^2 & \left(10^{-3}\right)^2 & \left(1.1 \cdot 10^{11}\right)^2 \end{bmatrix} \end{aligned} \tag{6.116}$$

$$R_k = \bar{w}_k^2 = 0.075^2 \tag{6.117}$$

The simulation results are as given in Figure 6.13.

As is seen by Figure 6.13, the estimation of the concentration of A is close to the true one (marginally better than for the additive noise model), but also in this case the estimated reaction rate and the estimated concentration of B suffer from noise. Also we are not able to explicitly tune the

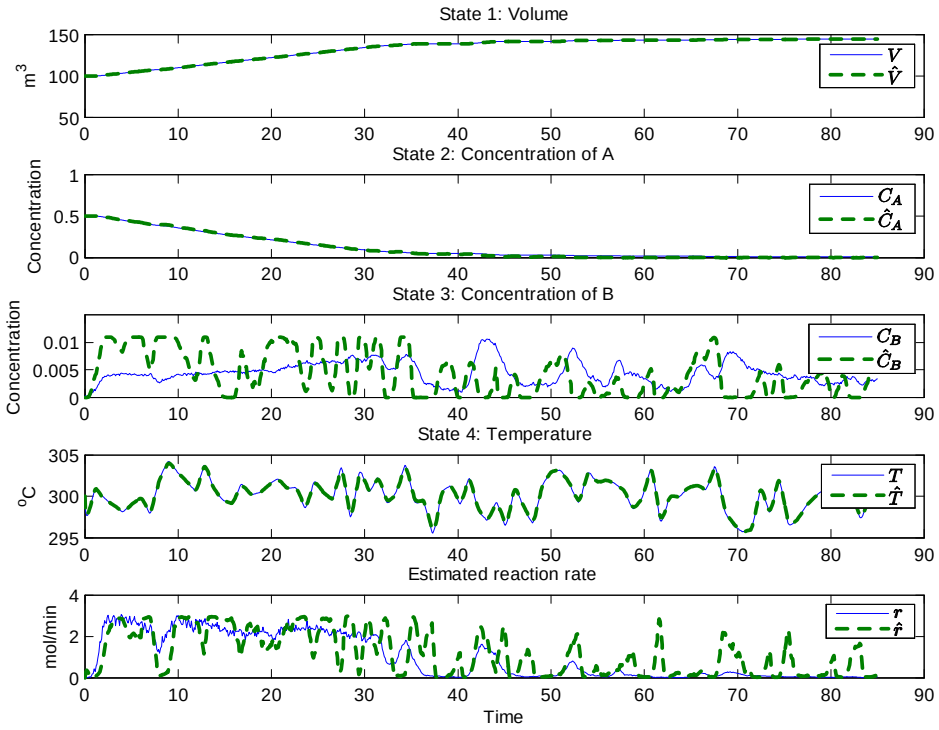

Figure 6.13: The results of the simulation assumed simplified model with noise entering the reaction rate $\hat{r}$.

behavior of the estimated concentration of B, since it is assumed that the noise enters via v_r. In the previous chapter we were able to relax the noise behavior in the estimated concentration of B by increasing the noise $\bar{v}_{x_3}$. The chosen noise model suffers in this case from the lack of this possibility.

Method 1, advanced model

Utilizing increased 'process knowledge' as given by (6.90), estimating k_0 instead of Q_r, the model is as given by (6.104)-(6.108) with the state vector as by (6.91). Further we erroneously assuming the standard approach that all the states experience additive noise, and no noise in the auxiliary parameters. The initial values are

$$\hat{x}_0 = \begin{bmatrix} 100 & 0.5 & 0 & 300 & 0.95k_0 \end{bmatrix}^T \tag{6.118}$$

$$\begin{array}{c} \textit{Sigma points lower bounds} : [0,\ 0,\ 0,\ 0,\ 8.5 \cdot 10^{11}]^T \\ \textit{Sigma points upper bounds} : [\infty,\ \infty,\ \infty,\ \infty,\ 9.5 \cdot 10^{11}]^T \end{array} \tag{6.119}$$

$$P_0 = diag\begin{bmatrix} \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(9.45 \cdot 10^{10}\right)^2 \end{bmatrix} \tag{6.120}$$

$$\begin{aligned} Qk &= diag\begin{bmatrix} \bar{v}_{k,x_1} & \bar{v}_{k,x_2} & \bar{v}_{k,x_3} & \bar{v}_{k,x_4} & \bar{v}_{k,x_5} \end{bmatrix} \\ &= diag\begin{bmatrix} \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(10^{11}\right)^2 \end{bmatrix} \end{aligned} \tag{6.121}$$

$$R_k = \bar{w}_k^2 = 0.075^2 \tag{6.122}$$

The results are as given by Figure 6.14.

As is seen by Figure 6.14, the estimation of all the states are close to the true ones, and the estimator also converges to the true state with respect to the estimated parameter k_0. However, we experienced that in the attempt to increase the convergence speed with respect to the estimated parameter k_0 (by increasing $\bar{v}_{x_5}$), the estimate of k_0 became more and more oscillatory.

Method 1 and 3 combined, advanced model

Utilizing extended 'process knowledge' as given by (6.90), estimating k_0 instead of Q_r, the model is as given by (6.104)-(6.108) with the state vector as by (6.91). H is as by (6.110) and HQH^T as by (6.111). It is assumed that the reaction rate (r) experience Gaussian noise and is given by

$$\hat{r} = \hat{k}_0 e^{\left(-\frac{ER}{\hat{x}_4}\right)} \hat{x}_2 \hat{x}_3^2 + v_r \tag{6.123}$$

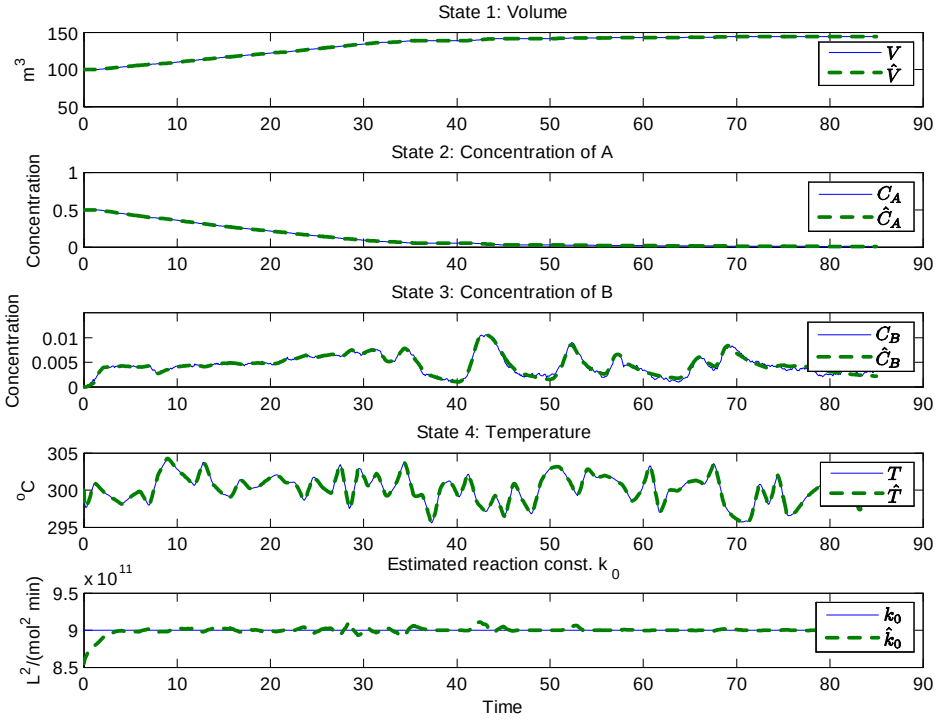

Figure 6.14: The results of the simulation assumed increased process knowledge and assuming additive noise.

The state vector is as given by (6.88) and the other initial values are

$$\hat{x}_0 = \begin{bmatrix} 100 & 0.5 & 0 & 300 & 0.95k_0 \end{bmatrix}^T \tag{6.124}$$

$$\begin{array}{c} \textit{Sigma points lower bounds} : [0,\ 0,\ 0,\ 0,\ 8.5 \cdot 10^{11}]^T \\ \textit{Sigma points upper bounds} : [\infty,\ \infty,\ \infty,\ \infty,\ 9.5 \cdot 10^{11}]^T \end{array} \tag{6.125}$$

$$P_0 = diag \begin{bmatrix} \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(10^{-5}\right)^2 & \left(9.45 \cdot 10^{10}\right)^2 \end{bmatrix} \tag{6.126}$$

$$\begin{aligned} Q_k &= diag \begin{bmatrix} \bar{v}_{k,x_1} & \bar{v}_{k,r} & \bar{v}_{k,x_5} \end{bmatrix} \\ &= diag \begin{bmatrix} \left(10^{-5}\right)^2 & \left(10^{-3}\right)^2 & \left(1.1 \cdot 10^{11}\right)^2 \end{bmatrix} \end{aligned} \tag{6.127}$$

$$R_k = \bar{w}_k^2 = 0.1^2 \tag{6.128}$$

The results are as given by Figure 6.15.

As is seen by Figure 6.15, the estimation of all the states are close to the true ones, and the estimator also converges to the true state with respect to the estimated parameter k_0. However, we experienced that in the attempt to further increase the convergence speed with respect to the estimated parameter k_0 (by increasing v_{x_5}), the estimate of k_0 become more and more oscillatory, but not as oscillatory as in the case of additive noise.

Some notes regarding Case '4 State CSTR'

By comparing our results with the results from Rao (2000) we see that our results differs. This is mainly because we have decreased the noise variance of the temperature measurement from 1 to 0.1. This can be justified by that measuring the temperature with 0.1^oC accuracy is not unusual. This is especially of great importance using the simplified model, but of minor importance using the advanced model.

6.6 Discussion

The results indicate that the noise modeling concepts based on input uncertainty has an effect. In both the '2 state CSTR' case and the industrial case the same modeling concept (combination of Method 1 and 2) is applied. However, the two simulated cases may be interpreted differently.

In the case of the '2 state CSTR' the purpose was to utilize 'process knowledge' that the states were uncertain when the control inputs were

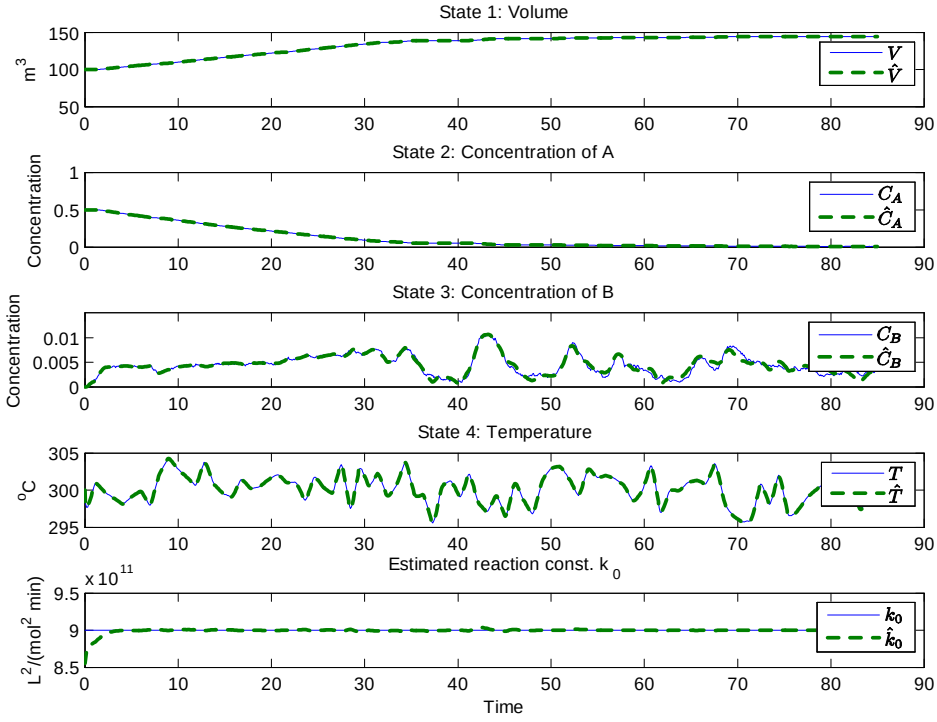

Figure 6.15: The results of the simulation assumed increased process knowledge and assuming noise entering the reaction rate $\hat{r}$.

applied to the process. This is reflected by injecting the uncertainty into the covariance. The common way to handle such a situation, if accounted for at all, is believed to be to reset the covariance $P(t)$ to its initial value $P(0)$.

In the industrial case there was actual uncertainty in the control input, and the purpose was indeed to let this knowledge be incorporated into the covariance. The traditional way to handle such a situation is to apply a parameter to the control input, say $k_1(t)u_1(t)$ and try to estimate k_1 (see e.g. Pannoncchia (2003)). The presented method may be seen as an alternative to the traditional way, but without the need to estimate the parameter k_1, and hence avoid the need for state vector augmentation.

The investigation of the noise entering auxiliary variables in the model, as covered by Method 3, shows that it might serve as an alternative method to explicit off-diagonal tuning of the "system noise matrix" Q/Q_k in cases where off-diagonal tuning is considered an option. Further, this method could be seen as a way not to jeopardize the mass/energy balance, in that the correlation between the states when it comes to how the uncertainty enters the states is fulfilled. This is seen by e.g. (6.46) and (6.47).

The investigation of the '4 state CSTR'-case showed that all the investigated noise models fulfilled the main challenge, namely to estimate the concentration of A. In the extended challenge, to accurately also estimate the reaction rate and the concentration of B, the choice of noise models made a significant difference. In the case of an simplified model, Method 3 did not improve the remaining estimates. In the case of applying a more accurate model, the use of Method 3 (combined with Method 1) showed marginally better estimation properties with respect to the estimated parameter k_0. Also, reducing the noise vector, implies fewer tuning parameters and noise parameters to be determined, which made the approach in the said case easier to tune than the additive noise approach.

6.7 Conclusion

We have presented some alternative noise modeling concepts which may serve as an effective alternative to the 'additive noise' paradigm.

Further, we have shown that letting noise entering auxiliary variables, it might serve as an alternative method to explicit off-diagonal tuning of the "system noise matrix" Q/Q_k in cases where off-diagonal tuning is considered an option.

6.8 Acknowledgements

The financial support of Hydro Aluminium AS is gratefully acknowledged. Also we acknowledge Dr. C. Rao for providing background information for the '4 state CSTR' case and professor J. B. Rawlings for the permission to use the flow data profile (F) in section 6.5.

Chapter 7

A recursive algorithm for computing dynamic eigenvalues

This chapter is based on (Kolås, Neerhoff, van der Kloet and Foss 2008) as submitted to Systems & Control Letters

7.1 Abstract

Calculating the so-called dynamic eigenvalues for linear time varying (LTV) systems involve solving non-symmetric differential Riccati equations (DRE). Solving these DRE's is challenging, especially when the solution involves singularities. In this paper we suggest a recursive computer algorithm for computing the dynamic eigenvalues based on the Möbius transform. Systems with periodic singularities are also considered.

Keywords: linear time varying systems, dynamic eigenvalues, Möbius transform

7.2 Introduction

For linear time invariant (LTI) systems it is well known that the eigenvalues of the system under investigation are well suited to analyze the stability directly. It is also well known that the use of the same technique for linear time varying (LTV) systems may fail (e.g. Wu (1974), Khalil (1996), Chen (1999)).

van der Kloet and Neerhoff (2000) have presented a theory for calculating the so-called dynamic eigenvalues for LTV systems. The dynamic eigenvalues cannot be used to determine the stability for LTV systems in a direct fashion. They are, however, closely related to the Lyapunov exponents (Neerhoff and van der Kloet (2000), van der Kloet and Neerhoff (2004)) which in their turn predicts the stability for LTV systems. In this work the complex Lyapunov exponent is considered (van der Kloet and Neerhoff (2004)).

Computing the dynamic eigenvalues by the use of a computer is not necessarily straightforward. In this paper we present a robust algorithm that computes the dynamic eigenvalues based on the Möbius transform (Schiff and Schnider (1999)).

The paper is organized as follows. First we recapture the eigenvalues for the LTI system and the flaw when used on LTV systems. Then we introduce the reader to the theory of dynamic eigenvalues. Next, we show how the Möbius transform can be used in computing the dynamic eigenvalues, present an algorithm based on the Möbius transform and show the results of the algorithm applied on some selected examples. Finally, we discuss some simulation results and point to some important issues regarding stability properties.

7.3 Eigenvalues for LTI-systems

Let the LTI system under investigation be defined by

$$\dot{x}(t) = Ax(t) \tag{7.1}$$

where A denotes the constant system matrix. The system is asymptotically stable if

$$\lim_{t\to\infty} x(t) = 0 \tag{7.2}$$

for all bounded initial states $x(0)$.

Equivalently, the system in (7.1) is asymptotically stable iff the real part of all the different eigenvalues of A are less than or equal to zero. If A is a 2x2 matrix, it can be shown that the eigenvalues are given by

$$\lambda_{1,2} = \frac{1}{2}\left(trace(A) \pm \sqrt{(trace(A))^2 - 4\det(A)}\right) \tag{7.3}$$

To motivate the use of another approach we use an example in which the classical approach from LTI systems is used on the LTV system

$$\dot{x}(t) = A(t)x(t) \tag{7.4}$$

where $A(t)$ denotes the time varying system matrix.

Example 1: Given the system (Chen (1999))

$$\dot{x}(t) = A(t)x(t) = \begin{bmatrix} -1 & e^{2t} \\ 0 & -1 \end{bmatrix} x(t) \tag{7.5}$$

From (7.3) we get the eigenvalues

$$\lambda_1 = \lambda_2 = -1 \quad \forall t \tag{7.6}$$

and the classical LTI approach predict that the system is stable for all t.

However, the solution of (7.5) is given by

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} e^{-t} & \frac{1}{2}(e^{t} - e^{-t}) \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} \tag{7.7}$$

and hence $x_1(t)$ will grow to infinity if $x_2(0)$ is different from zero, i.e. the system is not asymptotically stable.

The reason why the time varying properties of the matrix $A(t)$ is not transparent in the eigenvalues, is due to the off-diagonal elements, since the time varying e^{2t} is masked off in the determinant.

7.4 Dynamic Eigenvalues

LTV systems like (7.4) have been studied by a number of authors. The contribution by van der Kloet and Neerhoff can be seen as a generalization of the work presented in Wu (1980), Kamen (1988), Zhu and Johnson (1989), Zhu and Johnson (1991) and Zhu and Johnson (1997). In van der Kloet

and Neerhoff (2000) the concept of dynamic eigenvalues, calculated by the use of non-symmetric differential Riccati equations, are introduced.

First, the system matrix $A(t)$ in (7.4) is put into an upper triangular matrix $B(t)$ of same dimension. This step can be performed with a coordinate transformation

$$x(t) = R(t)y(t) \tag{7.8}$$

such that (7.4) is transformed into

$$\dot{y}(t) = B(t)y(t) \tag{7.9}$$

with

$$B(t) = R^{-1}(t)A(t)R(t) - R^{-1}\dot{R}(t) \tag{7.10}$$

Now, if

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix} \tag{7.11}$$

and

$$R(t) = \begin{bmatrix} 1 & 0 \\ p(t) & 1 \end{bmatrix} \tag{7.12}$$

then we will have

$$B(t) = \begin{bmatrix} \lambda_1(t) & a_{12}(t) \\ 0 & \lambda_2(t) \end{bmatrix} \tag{7.13}$$

with

$$\begin{aligned} \lambda_1(t) &= a_{11}(t) + p(t)a_{12}(t) \\ \lambda_2(t) &= a_{22}(t) - p(t)a_{12}(t) \end{aligned} \tag{7.14}$$

if and only if $p(t)$ satisfies the differential Riccati equation

$$\dot{p}(t) = -a_{12}(t)p^2(t) - (a_{11}(t) - a_{22}(t))p(t) + a_{21}(t) \tag{7.15}$$

The functions $\lambda_1(t)$ and $\lambda_2(t)$ are the eigenvalues of $B(t)$. Since (7.10) shows that $A(t)$ and $B(t)$ are not similar, the functions $\lambda_1(t)$ and $\lambda_2(t)$ are not the eigenvalues of $A(t)$. They are called the dynamic eigenvalues.

It can be shown that the solution of (7.9) can be written

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} e^{\gamma_1(t)} & 0 \\ 0 & e^{\gamma_2(t)} \end{bmatrix} \begin{bmatrix} 1 & f(t) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} \tag{7.16}$$

with

$$\gamma_i(t) = \int_0^t \lambda_i(\tau) d\tau \tag{7.17}$$

and

$$f(t) = \int_0^t a_{12}(\tau) e^{\gamma_2(\tau) - \gamma_1(\tau)} d\tau \tag{7.18}$$

Equation (7.16) shows that the asymptotic behavior of $[y_1(t) \;\; y_2(t)]^T$ is determined by $\gamma_1(t)$ and $\gamma_2(t)$. With (7.8) it is seen that this also is true for $[x_1(t) \;\; x_2(t)]^T$ if the transformation matrix $R(t)$ is bounded. Such a bounded coordinate transformation is called a Lyapunov transformation (Adrianova (1995)).

DEFINITION 1: Lyapunov transformation
The transformation (7.8) is called a Lyapunov transformation if

1. $R \in C^1[t_0, \infty >$
2. $R(t), R^{-1}(t), \dot{R}(t)$ *are bounded for* $t \geq t_0$.

The matrix $R(t)$ having these properties is said to be a Lyapunov matrix.

PROPERTIES 1: Notes the following properties of the Lyapunov transformations:

1. *Lyapunov transformations form a group.*
   Indeed
   (a) *if $R(t)$ is a Lyapunov matrix, then is also a $R^{-1}(t)$ Lyapunov matrix.*
   (b) *$R(t) \equiv I$ is a Lyapunov matrix.*
   (c) *if $R_1(t)$ and $R_2(t)$ are Lyapunov matrices, then $R(t) = R_1(t) \; R_2(t)$ is a Lyapunov matrix.*
2. *Lyapunov transformations do not change characteristic exponents* (for proofs, see Adrianova (1995)).

By *Properties 1* it is demonstrated that this class of Lyapunov transformations are an algebraic group with respect to multiplication.

Furthermore, if a transformation where the demands of *Definition 1* is met, the mean values turn out to be equal to the complex Lyapunov exponents (Neerhoff and van der Kloet (2000), van der Kloet and Neerhoff (2004)) as follows (see also Appedix B)

DEFINITION 2: Complex Lyapunov exponents

$$L_i = \lim_{t\to\infty} \frac{1}{t}\gamma_i(t) = \lim_{t\to\infty} \frac{1}{t}\int_0^t \lambda_i(\tau)d\tau \tag{7.19}$$

The constants L_i are known as complex Lyapunov exponents[1]. Note that it is obvious that $\gamma_i(t)$ for $t \to \infty$ is compared with $L_i t$ because

$$\lim_{t\to\infty} e^{\gamma_i(t)-L_i t} = 1 \tag{7.20}$$

It may be concluded that the solution $p(t)$ of the differential Riccati equation (7.15) determines the dynamic eigenvalues, which in turn determine via (7.8) and (7.16) the analytical solution of (7.4).

For an *n-th* order system the dynamic eigenvalues are found by partitioning the general matrix $A(t)$ (Wu (1984), van der Kloet and Neerhoff (2004), van der Kloet and Neerhoff (1999), van der Kloet and Neerhoff (2001), Neerhoff and van der Kloet (2001)) as

$$A(t) = \begin{bmatrix} A_{11}(t) & a_{12}(t) \\ a_{21}^T(t) & a_{nn}(t) \end{bmatrix} \tag{7.21}$$

and the matrix $R(t)$ by

$$R_n(t) = \begin{bmatrix} I_{n-1} & 0 \\ p_n^T(t) & 1 \end{bmatrix} \tag{7.22}$$

By this we get

$$B_n = \begin{bmatrix} A11(t) + a_{12}(t)p_n^T(t) & a_{12}(t) \\ 0^T & \lambda_n(t) \end{bmatrix} \tag{7.23}$$

with

[1] The Lyapunov exponent is to be interpretated as an average of the real part of a dynamic eigenvalue over a sufficient long period of time.

$$\lambda_n(t) = a_{nn}(t) - p_n^T(t)a_{12}(t) \tag{7.24}$$

if and only if $p^T(t)$ satisfies the non-symmetric vector differential Riccati equation

$$\dot{p}_n^T = -p_n^T a_{12} p_n^T - p_n^T A_{11} + a_{nn} p_n^T + a_{21}^T \tag{7.25}$$

To find the succeeding dynamic eigenvalue we have to look at the sub-matrix in (7.21) formed by

$$A_i = A_{11} + a_{12} p_n^T \tag{7.26}$$

and repeat the procedure from (7.21) on (7.26) with A_i in place for A in (7.21). For an nth order system the procedure is iterated $n-1$ times (nxn is the dimension of the system). The procedure will finally result in the upper triangular matrix $B(t)$ as

$$B(t) = \begin{bmatrix} \lambda_1(t) & \cdots & b_{ij}(t) \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n(t) \end{bmatrix} \tag{7.27}$$

To find the last eigenvalue one uses the fact that the procedure is trace preserving (Neerhoff and van der Kloet (2001)) in that

$$trace(A(t)) = trace(B(t)) = \sum_{i=1}^{n} \lambda_n(t) = \lambda_1(t) + \sum_{i=2}^{n} \lambda_n(t) \tag{7.28}$$

The compound transformation matrix $R(t)$ can be written as

$$R(t) = \begin{bmatrix} R_2 & 0 \\ 0 & I_{n-2} \end{bmatrix} \begin{bmatrix} R_3 & 0 \\ 0 & I_{n-3} \end{bmatrix} \cdots \begin{bmatrix} R_{n-1} & 0 \\ 0 & 1 \end{bmatrix} R_n \tag{7.29}$$

where

$$R_{n-i} = \begin{bmatrix} I_{n-i-1} & 0 \\ p_{n-i}^T & 1 \end{bmatrix} \tag{7.30}$$

and every p_{n-i}^Thas to satisfy the differential Riccati equation (7.25).

Further we note (see van der Kloet and Neerhoff (2000) and van der Kloet et al. (2007)) that the real part of the complex Lyapunov exponents, shortly known as the Lyapunov exponents, are given by

$$\chi_i = \mathrm{Re}(L_i) \tag{7.31}$$

If $\chi_i \leq 0$ the system is asymptotically stable, and if $\chi_i > 0$ the system is unstable if the dynamic eigenvalues and the Lyapunov exponents are preserved under the Lyapunov transformation (ref. *Definition 1* and *Properties 1*).

An example of using the theory is given below.

Example 2: Recapture the system given in (7.5)

$$\dot{x}(t) = A(t)x(t) = \begin{bmatrix} -1 & e^{2t} \\ 0 & -1 \end{bmatrix} x(t) \tag{7.32}$$

Using the theory of dynamic eigenvalues on (7.5) we obtain

$$A_{11} = -1,\ a_{12} = e^{2t},\ a_{21}^T = 0,\ a_{nn} = -1$$

From (7.25) we arrive at

$$\dot{p}(t) = -p^2(t)e^{2t} \tag{7.33}$$

which by separation of variables yields the solution

$$p(t) = \frac{1}{\frac{1}{p(t_0)} + \frac{1}{2}(e^{2t} - e^{2t_0})} = \frac{2p(t_0)}{2 + p(t_0)(e^{2t} - e^{2t_0})} \tag{7.34}$$

By (7.24) we get the dynamic eigenvalue

$$\lambda_2(t) = -1 - 2\frac{p(t_0)}{p(t_0) + (2 - p(t_0)e^{2t})e^{-2t}} \tag{7.35}$$

Assuming $t_0 = 0$, (7.35) becomes

$$\lambda_2(t) = -1 - 2\frac{p(0)}{p(0) + (2 - p(0))e^{-2t}}. \tag{7.36}$$

Inspection of (7.36) gives

$$\lambda_2(0) = -1 - 2\frac{p(0)}{p(0) + (2 - p(0))} = -1 - p(0) \tag{7.37}$$

$$\lambda_2(\infty) = -1 - 2 = -3 \tag{7.38}$$

Further, $\lambda_1(t)$ follows as

$$\lambda_1(t) = trace(A(t) - \lambda_2(t)) \tag{7.39}$$

From (7.36) and (7.39) we get the last dynamic eigenvalue

$$\lambda_1(t) = trace(A(t) - \lambda_2(t)) = -2 - (-1 - 2\frac{p(0)}{p(0) + (2 - p(0))e^{-2t}}) \tag{7.40}$$

By (7.39) and (7.40) it is clear that the Lyapunov exponents are given by

$$(\chi_1, \chi_2) = (1, -3) \tag{7.41}$$

which in this case indicates that the system is unstable, where the traditional concept of classical (or 'static') eigenvalue analysis fails (see (7.6)).

7.5 The Möbius transform

The concept of dynamic eigenvalues requires the solution of the non-symmetric vector differential Riccati equation (7.25). It is well known that Riccati equations can have finite escape times[2] when singularities takes place, so special attention to the numerical integration scheme is required.

An illustrative example is that the solution of the time varying harmonic oscillator equation

$$\dot{p}(t) = -p^2(t) - 1 \tag{7.42}$$

is given by

$$p(t) = -\tan(t) \tag{7.43}$$

[2]Finite escape time: The state of an unstable nonlinear system can go to infinity in finite time.

which is observed having a periodic behavior (see also Example 3 below). Given a numerical algorithm integrating (7.25) by Euler forward method, assuming no special actions taken, would never pass the first singular point, since for example $p(\pi/2) = -\infty$. This is shown in Figure 7.1.

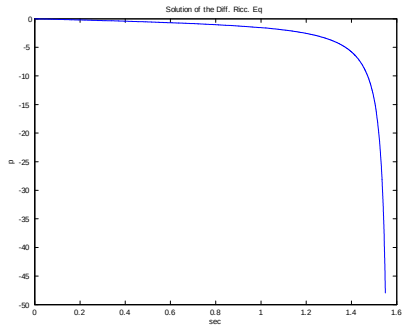

Figure 7.1: The figure illustrates that integrating (7.25) with Euler forward scheme applied on (7.42) leads to the solution $-Inf$ as time approach the first singular point. An algorithm based on Euler forward integration scheme would not pass this singular point.

Dealing with singularities when solving the differential Riccati equation numerically, has been extensively discussed in several papers (see Freiling (2002), Dieci (1992) and Schiff and Schnider (1999) for further references).

Schiff and Schnider (1999) shows that by considering the matrix

$$\Omega(t) = \begin{bmatrix} A_M(t) & B_M(t) \\ C_M(t) & D_M(t) \end{bmatrix} \tag{7.44}$$

with the associated differential Riccati equation

$$\dot{y}(t) = A_M(t)y(t) + B_M(t) - y(t)C_M(t)y(t) - y(t)D_M(t) \tag{7.45}$$

where $A_M(t) \in R^{kxk}$, $B_M(t) \in R^{kxl}$, $C_M(t) \in R^{lxk}$, $D_M(t) \in R^{lxl}$, the initial value problem of (7.45) is effectively integrated even through singularities via the explicitly one step Möbius-scheme of the form

$$y_{j+1} = \left(\tilde{A}_M(t_j, h)y_j + \tilde{B}_M(t_j, h)\right)\left(\tilde{C}_M(t_j, h)y_j + \tilde{D}_M(t_j, h)\right)^{-1} \tag{7.46}$$

In (7.46) the functions $\tilde{A}_M(t)$,$\tilde{B}_M(t)$, $\tilde{C}_M(t)$, $\tilde{D}_M(t)$ are constructed from the coefficient functions $A_M(t)$,$B_M(t)$, $C_M(t)$, $D_M(t)$ by a formula of the form

$$\tilde{\Gamma}_1(t_j,h) = \begin{bmatrix} \tilde{A}_M(t_j,h) & \tilde{B}_M(t_j,h) \\ \tilde{C}_M(t_j,h) & \tilde{D}_M(t_j,h) \end{bmatrix} = I_{k+l} + h\ \Omega(t) \tag{7.47}$$

or

$$\tilde{\Gamma}_1(t_j,h) = \begin{bmatrix} \tilde{A}_M(t_j,h) & \tilde{B}_M(t_j,h) \\ \tilde{C}_M(t_j,h) & \tilde{D}_M(t_j,h) \end{bmatrix} = I_{k+l} + h\ \Omega(t+\frac{h}{2}) + \frac{h^2}{2}\ \Omega(t+\frac{h}{2})^2 \tag{7.48}$$

By comparing (7.21) by (7.44), and (7.25) by (7.45) it is clear that if

$$\Omega(t) = \begin{bmatrix} A_M(t) & B_M(t) \\ C_M(t) & D_M(t) \end{bmatrix} = \begin{bmatrix} a_{nn} & a_{21}^T \\ a_{12} & A_{11} \end{bmatrix} \tag{7.49}$$

then $\dot{y} = \dot{p}^T$ and (7.45) becomes (7.25) and

$$p_{j+1}^T = \left(\tilde{A}_M(t_j,h)p_j^T + \tilde{B}_M(t_j,h)\right)\left(\tilde{C}_M(t_j,h)p_j^T + \tilde{D}_M(t_j,h)\right)^{-1} \tag{7.50}$$

By this fact we can use the Möbius transform to numerically solve the differential Riccati equation when computing the dynamic eigenvalues.

7.6 Algorithms for computing the Dynamic Eigenvalues

While the dynamic eigenvalues are relatively easy to find analytically for a 2x2 system, it is not necessarily so for larger systems. A 3x3-system could in fact be quite challenging. This motivates us to develop an algorithm of the concept.

By looking at the structure of the concept it is obvious that the total number of non-symmetric differential Riccati equations that has to be solved is given by

$$NoRiccati = \sum_{i=1}^{n-1} i \tag{7.51}$$

I.e. for $n = 5$ we have to solve 4 in the first iteration to compute λ_5, 3 in the next to compute λ_4, then 2 to compute λ_3 and finally 1 to compute λ_2– totally 10.

Developing an efficient algorithm requires some sort of splitting regarding the time. We may choose to integrate from t_0 to t_j in each iteration (later referred to as Algorithm 1), or try to make a recursive algorithm (later referred to as Algorithm 2) when it comes to the solutions of the differential Riccati equations ((7.25)) (see figure 7.2).

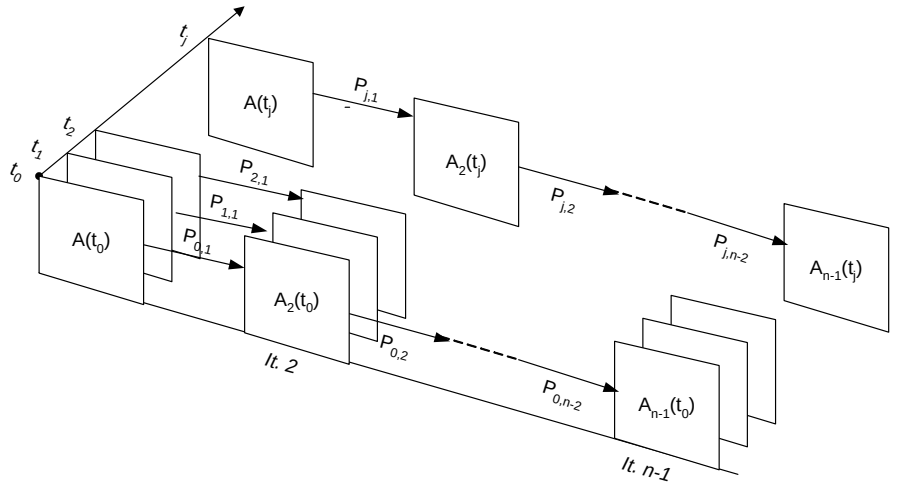

Figure 7.2: An illustration of the algorithm for computing the Dynamic Eigenvalues.

For a 2x2 system, the strategy to integrate from t_0 to t_j in an iteration $n-1$ (we have only one iteration in a 2x2 system) is doable. But if the integration window under investigation $(t_j - t_0)$ is large, or for small time steps the computation time may become very long. Also for larger systems the organization of the matrices and solutions of the differential Riccati equations, may become quite complex.

An alternative way is to discretize the differential Riccati equations (7.25) by the Möbius transform (7.50). This makes the algorithm recursive where the dynamic eigenvalues at time t_j becomes

$$\lambda_n(t_j) = a_{nn}(t_j) - p_j^T(t_j)a_{12}(t_j) \tag{7.52}$$

The Pseudo Code for Algorithm 2 is shown in Table 7.1.

Table 7.1: The Pseudo Code for the Möbius based Algorithm 2

```
Initialize:
p(0,:)=Initialize starting values Riccati

Body:
for jj=0 to j (see fig(7.2))
      for i=1 to n-1 (see fig(7.2))
         A(jj,i,:)= Partition the A-matrix as by (7.21) & (7.26)
         p(jj,i:)=SolveRiccatiByMöbius(A(jj,i,:)) as by (7.50) & (7.51)
         λ(jj, i, :) = Calculate dynamic eigenvalues as by (7.53)
         λ(jj, i, :) = Constrain dynamic eigenvalues(λ(jj, i, :))
      end
end
Evaluate the Lyapunov exponents according to (7.31)
```

7.7 Results

7.7.1 Algorithm efficiency

All simulations and development of the algorithms was done in Matlab Release 14. For Algorithm 1 the ode-solver 'ode45' was used, with relative tolerance and absolute tolerance set to 10^{-12}. The Möbius based Algorithm 2 was entirely coded in Matlab. The computer used was a Dell Precision M70 with Intel Pentium M processor running at 2.13Ghz and 2GB RAM.

To compare the performance of the algorithms we measured the time to run Example 2 with the system data given by (7.5) for $t \in [t_0, t_j] = [0, 5]$, $h = 0.005$, $p(t_0) = 3$, $x(t_0) = [1 \quad 1]^T$, and compared the solution of the differential Riccati equation with the analytical one ((7.34)). The results are shown in Table 7.2 and Figure 7.3 below.

Table 7.2: Comparison of the algorithms

Algorithm	Time used [sec]	$(\lambda_1(t_j), \lambda_2(t_j))$	$p_{analytical}(t_j) - p_{algorithm}(t_j)$
1	230	$(-3.0000, 1.0000)$	$-1.59 \cdot 10^{-13}$
2	0.9	$(-3.0000, 1.0000)$	$9.133 \cdot 10^{-7}$
Analytical	-	$(-3.0000, 1.0000)$	-

Figure 7.3 indicates that Algorithm 1 is more accurate than the Möbius based Algorithm 2, although the solution of the differential Riccati equation in Algorithm 2 converges close to the analytical solution. This is indicated

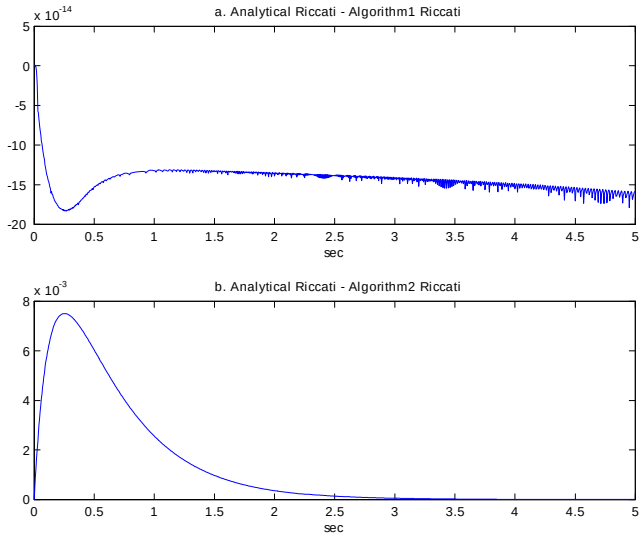

Figure 7.3: The uppermost plot (a) shows the difference between the analytical solution to the differential Riccati equation (25) and the solution found by Algorithm 1. The lowermost plot (b) shows the difference between the analytical solution to the differential Riccati equation (25) and the solution found by the Möbius based Algorithm 2.

by the last column in Table 7.2. However, as Table 7.2 in this particular case indicates, Algorithm 2 is approximately 250 times faster than Algorithm 1, but at the expense of some loss of accuracy. If the loss in accuracy is not acceptable, a more accurate Möbius approximation may be used (see (7.48) or Schiff and Schnider (1999) for higher order approximations).

Note also that Algorithm 1 fails when used on systems with periodic singularities, due to the integration method used.

7.7.2 Simulation results

The algorithms are tested on several cases found in the literature (see the Appendix regard Algorithm 2), but it may be of special interest to see how the algorithms works on periodic systems.

Below, an example is given, giving both the analytical and the result of the Möbius based algorithm (Algorithm 2). Note that Algorithm 1 fails on this example, because the integration algorithm does not survive the first singular point due to the periodicity of the system.

Example 3: Consider the mass/spring-system described by

$$\begin{array}{c} m\ddot{y}(t) + k\ y(t) = 0 \\ \Updownarrow \\ \ddot{y}(t) + q\ y(t) = 0, \qquad q = \frac{k}{m} \end{array} \tag{7.53}$$

By choosing $x_1 = y$, $x_2 = x_1$ and $x = [x_1 \ \ x_2]^T$, we get the system

$$\dot{x}(t) = A(t)x(t) = \begin{bmatrix} 0 & 1 \\ -q & 0 \end{bmatrix} x(t) \tag{7.54}$$

It can be shown that this system is Lyapunov stable when $q > 0$ and unstable if $q \leq 0$. The differential Riccati equation to be solved is

$$\dot{p}(t) = -p^2(t) - q \tag{7.55}$$

and it can be shown that the solution of the differential Riccati equation (7.55) in the case of $q > 0$ and $p(0) = 0$ is

$$p(t) = -\sqrt{q}\tan(\sqrt{q}t) \tag{7.56}$$

The dynamic eigenvalues are

$$(\lambda_1, \lambda_2) = (\sqrt{q}\tan(\sqrt{q}t), -\sqrt{q}\tan(\sqrt{q}t)) \tag{7.57}$$

I.e. $p(t)$ is periodic and bounded, except for a countable infinite set of isolated points with measure zero (stable oscillator).

Further, what is interesting in this case, is to investigate wether or not the proposed Möbius based algorithm survives the singularities, reflects the periodic solution of the differential Riccati equation and to investigate whether or not the Lyapunov exponents could tell something about the stability of this system.

By selecting $q = 1$, $h = \pi/50000$, $t = 0 - 10\pi$, $p(0) = 0$ we get the results (Algorithm 2) as shown in figure (7.4).

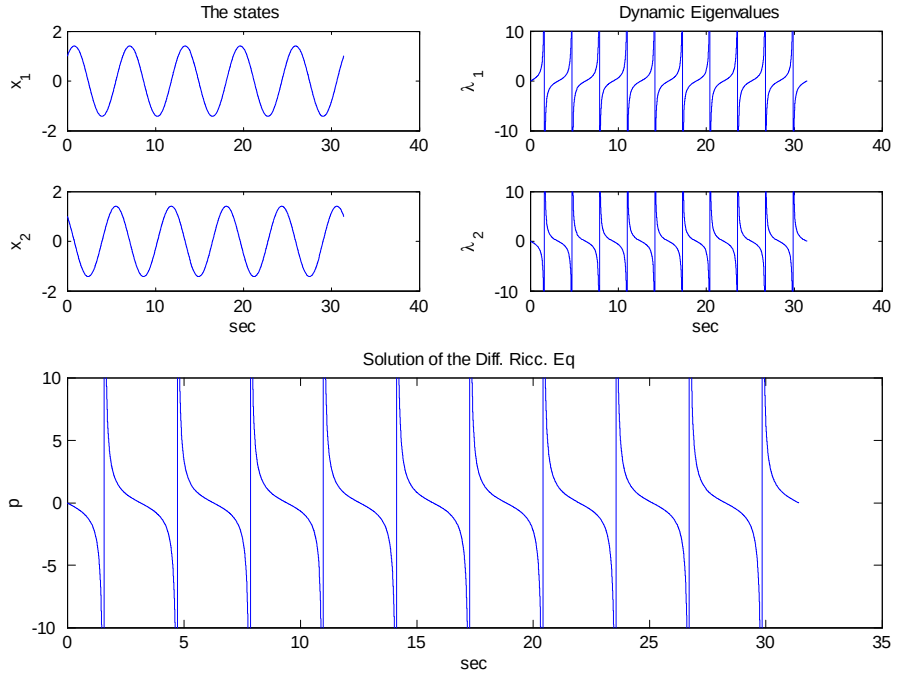

Figure 7.4: The results of the Möbius algorithm (Algorithm 2) applied to the case described by Example 3.

The Lyapunov exponents for $-10 \leq \lambda(t) \leq 10$ yields

$$(\chi_1, \chi_2) = (0.0002, -0.0002) \tag{7.58}$$

and we may conclude that the Lyapunov exponents indicates that the system is Lyapunov stable. Strictly speaking, the results of (7.58) indicates an unstable system, but it can be shown that the results regarding (7.58)

is influenced by the integration length, time step h and the order of the Taylor series used in the Möbius approximation (ref. (7.47) or (7.48)). By truncating the results of (7.58) after the 3^{rd} decimal, it may be reasonable to conclude that the results indicates a stable system.

It should be clear from the simulations shown in figure (7.4), that the proposed Algorithm 2 survives the singularities, and reflects the periodic solutions of (7.56) and (7.57).

Also we recognize the following in the case of periodic singularities with respect to Algorithm 2: The algorithm do indeed pass the singularities, but due to possible mismatch between the selection of the time step h and the periodicity of the system under investigation, the value of the differential Riccati equation may vary close to the singularities (see e.g. figure (7.7) at page 199 in the Appendix), although the solution is periodic. This will influence the accuracy of the dynamic eigenvalues and hence the evaluation of the Lyapunov exponents, which is based on mean value calculations. Because of this, constraining the values of the dynamic eigenvalues before calculating the Lyapunov exponents may be necessary.

Instability

We have run Algorithm 2 on stable and unstable LTI and LTV systems from dimension 2 to 10, and have found that the algorithm correctly identify both stable and unstable systems. However this was on systems where the matrix $R(t)$ was Lyapunov according to *Definition 1*, i.e. the solution of the differential Riccati equation was bounded, or only violates the Lyapunov definition with some solutions with measure zero (periodic solutions).

One should be aware of that the Lyapunov exponent as defined in *Definition 2* only predicts the stability for the system under investigation if the Lyapunov transformation meets the demands of *Definition 1*. If the Lyapunov transformation does not meet the demands of *Definition 1*, the suggested approach should not be applied. This is illustrated by the following examples.

Numerical issues: The following example addresses a special numerical behavior that may occur in using the algorithms.

Example 4: Given

$$\dot{x}(t) = A(t)x(t) = \begin{bmatrix} -1 & 0 \\ 0.1 & -1 \end{bmatrix} x(t) \tag{7.59}$$

Then

$$A_{11} = -1,\ a_{12} = 0,\ a_{21}^T = 0.1,\ a_{nn} = -1 \tag{7.60}$$

and the differential Riccati equation (7.25) becomes

$$\dot{p}(t) = 0.1 \tag{7.61}$$

which yields the solution

$$p(t) = p(t_0) + 0.1(t - t_0) \tag{7.62}$$

Assuming $t_0 = 0$ we get

$$p(t) = p(0) + 0.1t \quad \forall t \tag{7.63}$$

and by (7.24) the dynamic eigenvalue is found to be

$$\lambda_2(t) = a_{nn} - p(t)a_{12} = -1 - p(t) \cdot 0 = -1, \quad \forall t \tag{7.64}$$

By (7.28) we get the last dynamic eigenvalue

$$\lambda_1(t) = trace(A(t) - \lambda_2(t)) = -2 - (-1) = -1, \quad \forall t \tag{7.65}$$

Apparently the approach gives the correct answer with respect to predicting the stability, but as can be seen by (7.63), the solution of the differential Riccati equation grows to infinity, which means that $R(t)$ is not bounded (not Lyapunov). This may cause internal numerical problems in the proposed algorithms.

System representation issues: The following example addresses challenges due to system representation.

Example 5: Given the system

$$\dot{x}(t) = A(t)x(t) = \begin{bmatrix} -1 & 0 \\ e^{2t} & -1 \end{bmatrix} x(t) \tag{7.66}$$

This system is the conjugate to the system given in Example 1, as can be seen by using the transformation

$$x(t) = \Gamma\ z(t) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} z(t) \tag{7.67}$$

which related to (7.66) becomes

$$\dot{z}(t) = \Gamma^{-1} A(t)\ \Gamma\ z(t)\ = \begin{bmatrix} -1 & e^{2t} \\ 0 & -1 \end{bmatrix} z(t) \tag{7.68}$$

As a consequence, the solution of (7.66) can be written as

$$x(t) = \begin{bmatrix} e^{-t} & 0 \\ \frac{1}{2}\left(e^{t} - e^{-t}\right) & e^{-t} \end{bmatrix} x(0) \tag{7.69}$$

The matrix in (7.69) is simply the conjugate of the matrix in (7.7). Moreover the eigenvalues of the matrix in (7.66) are given by (7.6). Again we see that the eigenvalues of $A(t)$ have no meaning for determining stability.

To obtain a triangular form for the system in (7.66) the transformation ((7.12))

$$x(t) = R(t)y(t) = \begin{bmatrix} 1 & 0 \\ p(t) & 1 \end{bmatrix} y(t) \tag{7.70}$$

is applied.

This yields for (7.13) in the case

$$\dot{y}(t) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} y(t) \tag{7.71}$$

if and only if

$$\dot{p}(t) = e^{2t} \tag{7.72}$$

Thus for example if

$$p(t) = \frac{1}{2} e^{2t} \tag{7.73}$$

it is easy to see that the transformation matrix in this case is

$$R(t) = \begin{bmatrix} 1 & 0 \\ \frac{1}{2}e^{2t} & 1 \end{bmatrix} \tag{7.74}$$

This transformation is not bounded and thus it does not meet the demands of *Definition 1.* Thus (7.74) is not a Lyapunov transformation, and the characteristic exponents of the system and the dynamic eigenvalues may change with such a non-Lyapunov transformation. Hence, the suggested approach could not be used on this system.

To determine if the approach given in this paper can be used, one must look at the *Definition 1* and *Properties 1*: If $R(t)$ is bounded then the

concept of dynamic eigenvalues can be used to determine the stability of the investigated system, since then the Lyapunov exponents and hence the dynamic eigenvalues are preserved. This fact can be utilized in a numerical algorithm calculating the dynamic eigenvalues.

7.8 Conclusions

We have in this paper proposed two algorithms that calculates the dynamic eigenvalues based on the solutions of differential Riccati equations.

Particularly we have demonstrated that the proposed Möbius based algorithm (Algorithm 2) computes the dynamic eigenvalues close to the analytical solution for both stable and non-stable systems. In this respect we also recognize that the proposed Algorithm 2 computes the dynamic eigenvalues also for periodic systems, since singular points are effectively dealt with in the Möbius transform.

We also recognize that for the tested periodic systems, the Lyapunov exponents as defined by *Definition 2* predicts the stability properties for all tested cases, even though $R(t)$ is not bounded due to the periodic singularities.

Further we have pointed to some limitation in the proposed algorithms due to equidistant integration time. A natural extension of this work would be to implement the algorithm with variable time step. Another area of investigation would be to look at how one could reduce the influence of the singularities in the Lyapunov exponents in the case of periodic solutions.

Although we have demonstrated that our approach identifies the stability properties for several different systems, we have also pointed to the fact that the Lyapunov exponent as defined in *Definition 2* predicts the stability for the system under investigation only if the Lyapunov transformation meets the demands of *Definition 1.* A natural extension of this work would be to add to the algorithm the possibility to also predict the stability of systems where the transformation is not Lyapunov, and by that extending the class of LTV systems that can be analyzed with respect to stability issues. However, that includes the evaluation of the dynamic eigenpairs (dynamic eigenvectors and eigenvalues, see Appedix B and ((van der Kloet and Neerhoff, 2004, Chapter 2))), and is outside the scope of this paper.

7.9 Acknowledgement

The financial support of Hydro Aluminium AS is gratefully acknowledged. The authors acknowledge the following persons for the support and input during the writing of this paper: Dr. H. Siahaan and Dr. R. Kandepu at NTNU.

7.10 Appendix A - Tested Cases

The proposed algorithm (Algorithm 2) is tested on a wide range of cases. In this Appendix we list some of the cases tested together with the results. In all simulations we used a Möbius transform based on (7.47), and all the system equations were integrated by using the Central Difference integration scheme.

7.10.1 Case 1

System (Khalil, 1996, p.158):

Given the system

$$\dot{x}(t) = \begin{bmatrix} -1 + 1.5\cos^2(t) & 1 - 1.5\sin(t) \\ -1 + 1.5\sin(t)\cos(t) & -1 + 1.5\sin^2(t) \end{bmatrix} x(t) \tag{7.75}$$

$$x(0) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T,\ t \in [0, 50],\ p(t_0) = 0,\ h = 0.005 \tag{7.76}$$

Simulation results

The simulation results for Case 1 is shown in Figure(7.5).

The Lyapunov exponents for $-10 \leq \tau(t) \leq 10$ yields

$$(\chi_1, \chi_2) = \left(\bar{\lambda}_1, \bar{\lambda}_2\right) = (-0.9, 0.5) \tag{7.77}$$

and thus indicates this as an unstable system.

Using the classical LTI approach, the real part of the eigenvalues is

$$(\lambda_1, \lambda_2) = (-0.25, -0.25) \tag{7.78}$$

and thus identifies this system as stable.

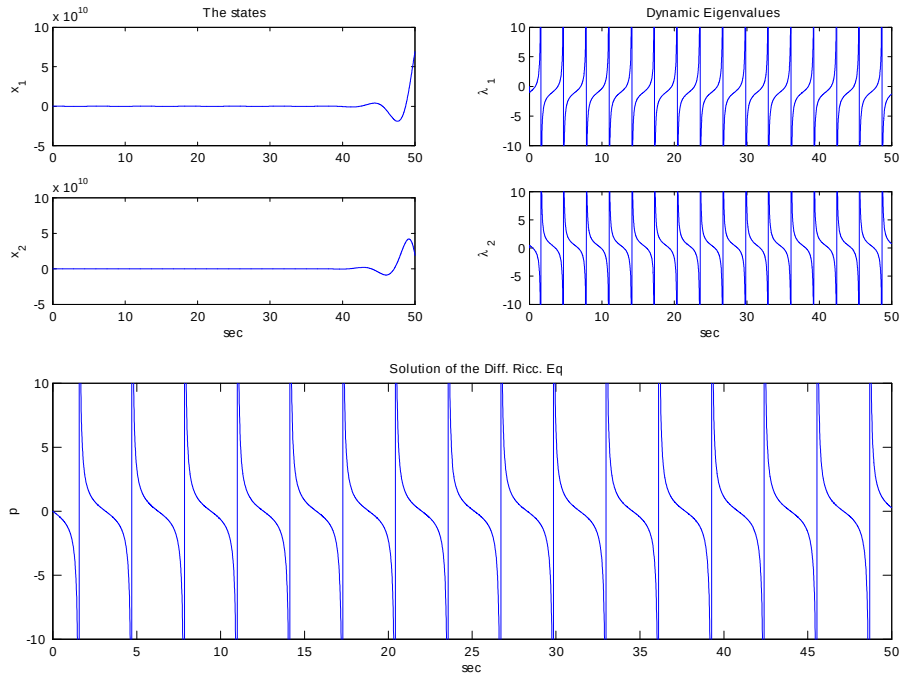

Figure 7.5: The results of the Möbius algorithm (Algorithm 2) applied to the case described by Case 1.

7.10.2 Case 2

System (modified version of van der Kloet and Neerhoff (2004))

Given the system

$$\dot{x}(t) = \begin{bmatrix} 1/3 - 5/2\cos^2(t) & 1 + 5/2\sin(t)\cos(t) \\ -1 + 5/2\sin(t)\cos(t) & 1/3 - 5/2\sin^2(t) \end{bmatrix} x(t) \tag{7.79}$$

$$x(0) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T,\ t \in [0, 50],\ p(t_0) = 3,\ h = 0.005 \tag{7.80}$$

Simulation results

The simulation results for Case 2 is shown in Figure(7.6).

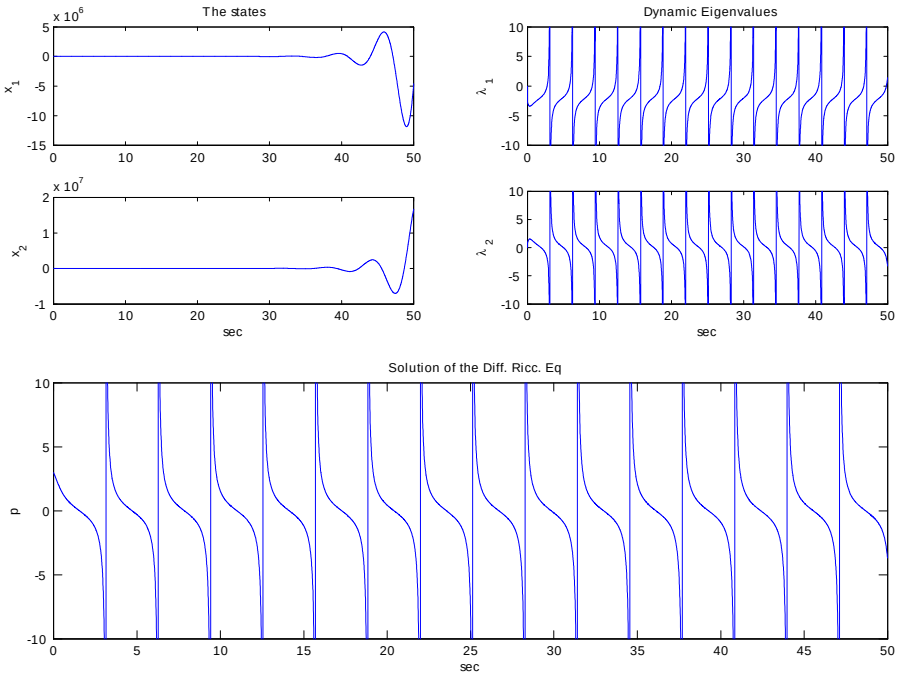

Figure 7.6: The results of the Möbius algorithm (Algorithm 2) applied to the case described by Case 2.

The Lyapunov exponents for $-10 \leq \tau(t) \leq 10$ yields

$$(\chi_1, \chi_2) = \left(\bar{\lambda}_1, \bar{\lambda}_2\right) = (-2, 0.3) \tag{7.81}$$

and thus indicates this as an unstable system.

Using the classical LTI approach, the eigenvalues becomes

$$(\lambda_1, \lambda_2) = (-1/6, -10/6) \tag{7.82}$$

and thus identifies this system as stable.

7.10.3 Case 3

System Wu (1974):

Given the system

$$\dot{x}(t) = \begin{bmatrix} -11/2 + 15/2\sin(12t) & 15/2\cos(12t) \\ 15/2\cos(12t) & -11/2 - 15/2\sin(1t) \end{bmatrix} x(t) \tag{7.83}$$

$$x(0) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T, t \in [0, 50], p(t_0) = 0, h = 0.005 \tag{7.84}$$

Simulation results

The simulation results for Case 3 is shown in Figure(7.7).

The Lyapunov exponents for $-100 \leq \tau(t) \leq 100$ yields

$$(\chi_1, \chi_2) = \left(\bar{\lambda}_1, \bar{\lambda}_2\right) = (-9.6, -1) \tag{7.85}$$

and thus indicates this as an stable system.

Using the classical LTI approach, the eigenvalues are

$$(\lambda_1, \lambda_2) = (2, -13) \tag{7.86}$$

and thus identifies this system as unstable.

7.10.4 Case 4

System Wu (1984):

Given the system

$$\dot{x}(t) = \begin{bmatrix} 2 & -e^{-t} \\ e^{-t} & 1 \end{bmatrix} x(t) \tag{7.87}$$

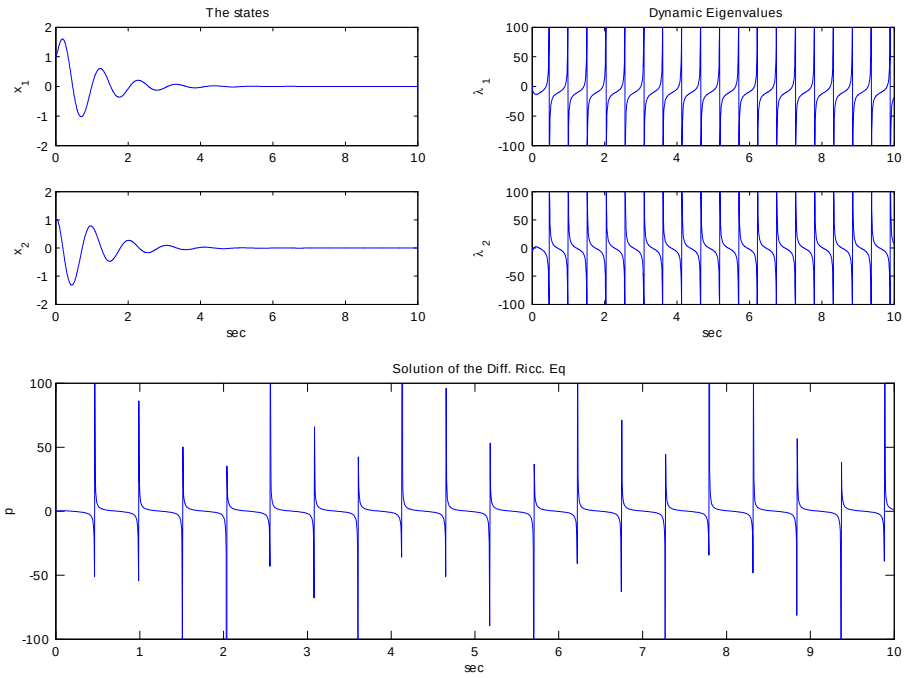

Figure 7.7: The results of the Möbius algorithm (Algorithm 2) applied to the case described by Case 3.

$$x(0) = [\ 1 \quad 1\ ]^T,\ t \in [0, 10],\ p(t_0) = 0,\ h = 0.001 \tag{7.88}$$

Simulation results

The simulation results for Case 4 is shown in Figure(7.8).

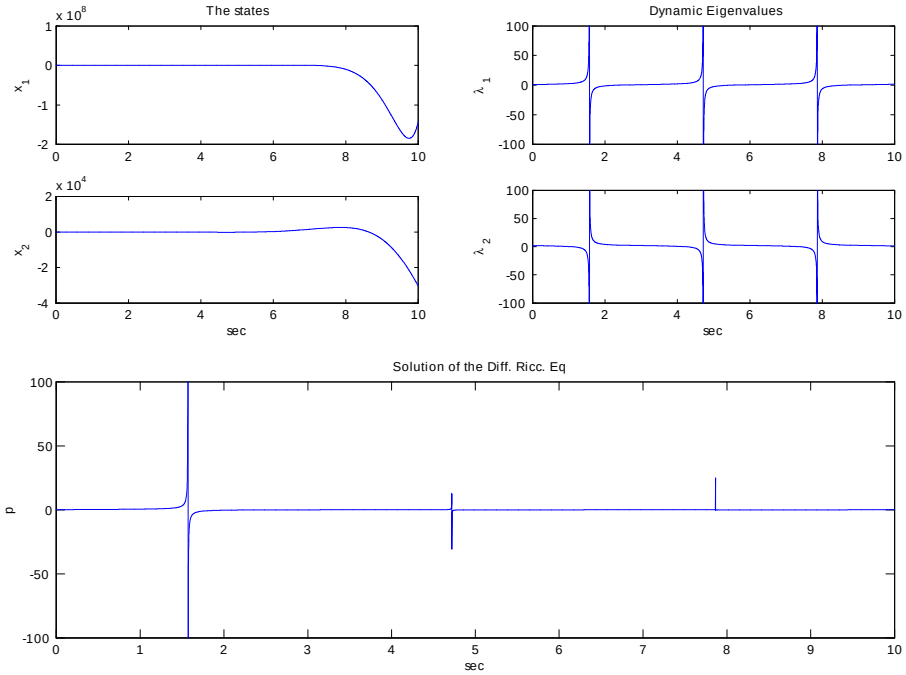

Figure 7.8: The results of the Möbius algorithm (Algorithm 2) applied to the case described by Case 4.

The Lyapunov exponents for $-100 \leq \tau(t) \leq 100$ yields

$$(\chi_1, \chi_2) = \left(\bar{\lambda}_1, \bar{\lambda}_2\right) = (1.1, 1.9) \tag{7.89}$$

and thus indicates this as an unstable system.

Using the classical LTI approach, the eigenvalues are

$$(\lambda_1, \lambda_2) = (1.5, 1.5) \tag{7.90}$$

and thus identifies this system as unstable.

7.10.5 Case 5

System:

Given the system

$$\dot{x}(t) = \begin{bmatrix} -1 & 1 & 1 & e^{4t} \\ 0 & -2 & 1 & 1 \\ 0 & 0 & -1.5 & 1 \\ 0 & 0 & 0 & -3 \end{bmatrix} x(t) \tag{7.91}$$

$$x(0) = [\ 1 \quad 1 \quad 1 \quad 1\ ]^T,\ t \in [0,6],\ p(t_0) = [\ 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\ ],\ h = 0.001 \tag{7.92}$$

Simulation results

The simulation results for Case 5 is shown in Figure(7.9)-(7.10).

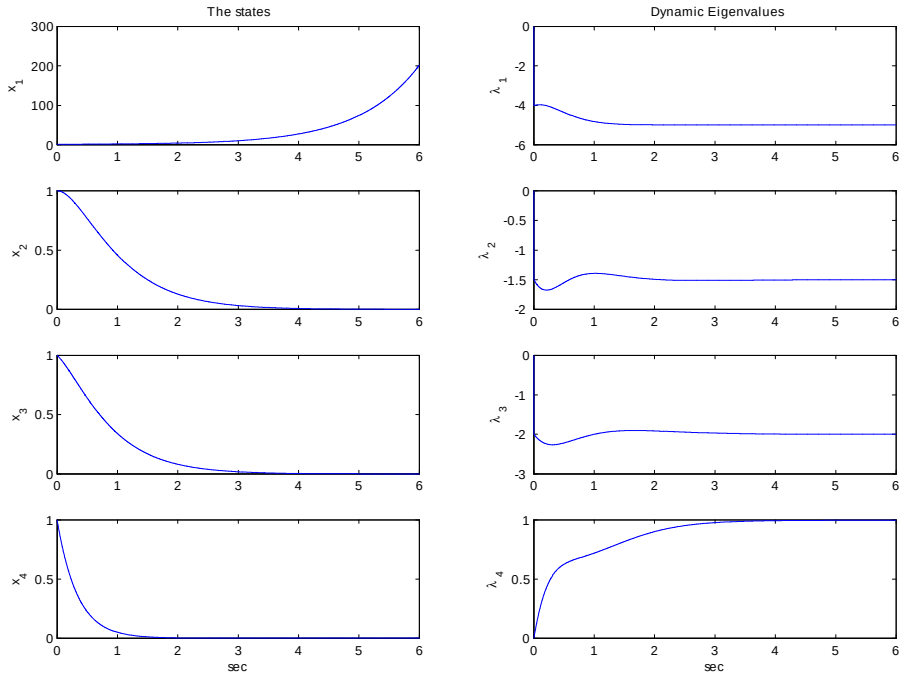

Figure 7.9: The results of the Möbius algorithm (Algorithm 2) applied to the case described by Case 5.

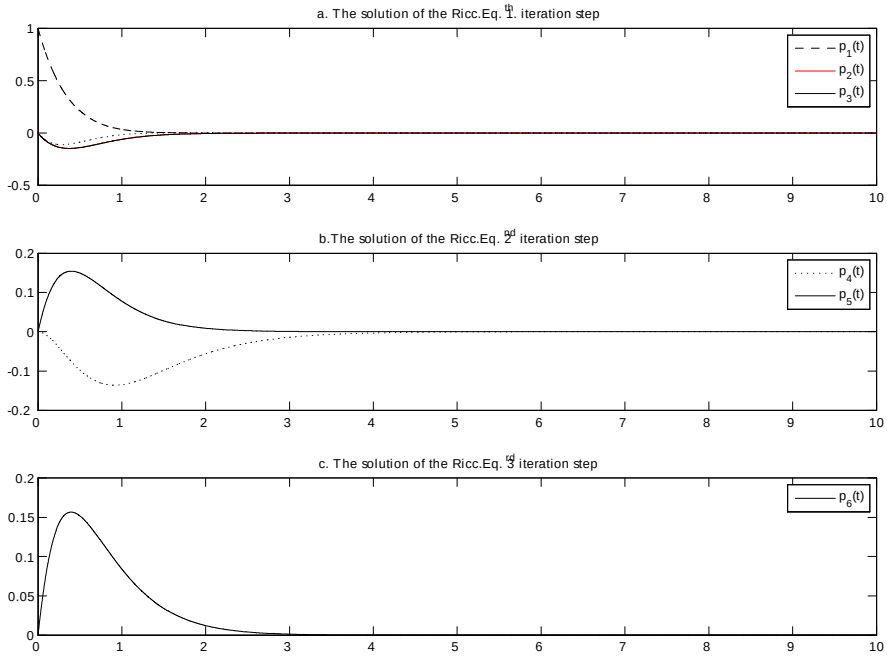

Figure 7.10: The solution of the Riccati Equations for Case 5. $p_1(t)$, $p_2(t)$ and $p_3(t)$ is the set of Riccati Equations for the first iteration step, $p_4(t)$ and $p_5(t)$ is the set for the second iteration step and, finally, $p_6(t)$ for the last iteration step.

The Lyapunov exponents yields

$$(\chi_1, \chi_2, \chi_3, \chi_4) = (-5, -1.5, -2, 1) \tag{7.93}$$

and thus indicates this as an unstable system.

Using the classical LTI approach, the real part of the eigenvalues are

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (-1, -2, -1.5, -3) \tag{7.94}$$

and thus identifies this system as stable.

7.10.6 Summary investigated systems

As a guide in getting an overview of the investigated systems throughout this paper with respect to the stability identification properties, a table is given below. We have used 1 to indicate that the method under investigation identified the system as stable and 0 as unstable.

Table 7.3: The table gives an overview of the investigated systems throughout this paper with respect to the stability identification properties of our apporach.

System	*Eigenvalues*	*Our approach*	*True system*	*Ref.page*
Example 2	1	0	0	182
Example 3	1	0	0	189
Example 4	1	1	1	191
Example 5	1	1	0	192
Case 1	1	0	0	195
Case 2	1	0	0	197
Case 3	0	1	1	198
Case 4	0	0	0	198
Case 5	1	0	0	201

As Table(7.3) indicates, the traditional eigenvalues applied on these cases fails in 7 of 9 cases, while our approach predicts the stability behavior in the cases where the transformation met the demands from *Definition 1.*

7.11 Appendix B - Eigenpairs

Let

$$\dot{x}(t) = A(t)x(t) \tag{7.95}$$
$$y(t) = R(t)x(t) \tag{7.96}$$

such that

$$\dot{y}(t) = \begin{bmatrix} \lambda_1(t) & a_{12}(t) \\ 0 & \lambda_2(t) \end{bmatrix} y(t) \tag{7.97}$$

Then R is a Riccati transformation. The solution implies

$$y(t) = \sum c_i \; e^{\gamma_i(t)} \tag{7.98}$$
$$c_i \; constant \tag{7.99}$$
$$e^{\gamma_i(t)} = \int_0^t \lambda_i(\tau)\, d\tau \tag{7.100}$$

This implies that the solution is the sum of the modal solutions

$$x(t) = R^{-1}(t)\; y(t) = \sum \left(R^{-1}(t)\; c_i\right)\; e^{\gamma_i(t)} = u_i(t) e^{\gamma_i(t)} \tag{7.101}$$

where $u_i(t)$ is the mode amplitude or eigenvector. From (Adrianova, 1995, eq (2.1.1)) we have the Lyapunov exponent defined as

$$\chi\left(x(t)\right) = \lim_{t\to\infty} \frac{1}{t} \ln |x(t)| \tag{7.102}$$

I.e. that any mode is equivalent with the Lyapunov exponent, which again is equivalent with the dynamic eigenpairs (i.e. dynamic eigenvalue $\lambda_i(t)$ + the dynamic eigenvector $u_i(t)$). That is

$$\chi\left(x(t)\right) = \lim_{t\to\infty} \frac{1}{t} \ln \left|R^{-1}(t)\; c_i\right| + \lim_{t\to\infty} \frac{1}{t}\; \gamma_i(t) \tag{7.103}$$
$$= \lim_{t\to\infty} \frac{1}{t} \ln |u_i(t)\;| + \lim_{t\to\infty} \frac{1}{t}\; \gamma_i(t) \tag{7.104}$$

Conclusion:

1. Dynamic eigenvalues λ_i are invariant under transformation $R(t)$ even if $R^{-1}(t)$ is not Lyapunov.

2. The mode stability is determined by both λ_i and u_i.

3. If $R^{-1}(t)$ bounded $\Longrightarrow$ $u_i(t)$ bounded $\Longrightarrow$ $R(t)$ is

   (a) a Lyapunov Transformation $\Longrightarrow \chi\left(x_i(t)\right) = \frac{1}{t}\,\gamma_i(t) = \int_0^t \lambda_i\left(\tau\right) d\tau$

   (b) and that Lyapunov exponents χ is also invariant under the transformation.

Note that 1. and 2. implies that if R is not Lyapunov, then also the dynamic eigenvectors have to be considered, and not only the dynamic eigenvalues.

Note also that *Definition 2* follows from 3a.

Part III

Contribution regarding knowledge processes

Chapter 8

The use of modeling and simulation to conserve knowledge

This chapter is not published. The chapter is written with contribution from associate professor R. Klev, NTNU, regarding Chapter 8.2-8.5.

8.1 Abstract

In an organization where knowledge is a key factor, is there a possibility to minimize the risk of knowledge disappearing with people when outsourcing, turnover or downsizing of the organization is an issue? Further, if the knowledge is complex how can an organization speed up time-consuming learning of employees without making every employee an expert? The use of modeling and simulation as design and construction tools is well known, but there has been little focus on the view of modeling and simulation seen in the context of conserving knowledge. The possibility that opens up with the use of modeling and simulation in the context of learning, knowledge creation and knowledge sharing is exciting. This chapter addresses these issues by analyzing the role of modeling and simulation in the context of conserving knowledge.

Keywords: modeling, simulation, learning, knowledge creation, knowledge sharing, conserving knowledge

8.2 Introduction

The continuous struggle for effectiveness and profitability, facing most industrial organizations, demand frequent structural changes in terms of reorganization, outsourcing or downsizing. This implies that many people will leave their positions in the organization, and important knowledge could be lost. At the same time, recent developments within strategic management and organizational theory argue on the crucial importance of individual and organizational knowledge for growth and development of companies (Hislop (2005), Carlsen et al. (2004), von Krogh et al. (2000), Nonaka and Takeuchi (1995), Spender (1993)). This leaves the company with the combined challenge of preserving and developing its knowledge resources, while at the same time allowing for cutting number of personnel involved in each work processes. In other words, the company has to make crucial knowledge and learning processes less dependent on individuals.

In this article, we will explore the role and possibilities of modeling and simulation in the context of learning, knowledge creating and the sharing of knowledge. Based on a discussion of the nature of knowledge and learning, the principles of modeling and simulation, and analysis of two practical cases, we suggest that developing mathematical models for computer simulations has a potential for being useful in at least three different forms of knowledge processes:

- It could enhance the organizations capacity to conserve/and or reproduce important knowledge, and thus make a knowledge dependent organization less vulnerable of people leaving their position.

- It could be a tool for knowledge sharing and knowledge creation between people with different areas of expertise, and

- It could be instrumental in creating and communication simplified and crucial insights into (in principle) processes with high complexity.

In order to substantiate this position, we will present some main positions within the field of knowledge and learning in organizations. Further; relevant topics of modeling and simulation are presented, followed by a discussion where modeling and simulation are put into the context of learning, knowledge creation and knowledge sharing.

8.3 Knowledge

There are two major epistemological camps in the debate on the nature of knowledge (Hislop (2005), Stenmark (2002); Klev and Carlsen (2000)): On the one side, there is the objectivist perspective, where the basic assumption is that knowledge can be treated similar to a tangible physical resource in the organization. On the other, we find the practice-based perspective which basically claims that knowledge is essentially personal and embedded in individual and organizational practices.

To understand the practical importance of the differences between positions, we should emphasize the discussion of tacit vs explicit knowledge (Nonaka and Takeuchi (1995)). Explicit knowledge is knowledge that can be codified and expressed in documents, diagrams and/or in a computer. In an objectivist position, knowledge is in principle possible to explicate. If it cannot be made explicit, it really isn't knowledge. Tacit knowledge (Polanyi (1983)) is the knowledge that is used effortlessly and routinely as part of work, but which is hard (or impossible) to articulate in an explicit form. From a practice-based perspective, the knowledge is essentially tacit and embedded within the cultural values and assumptions of those who possesses and uses it (Hislop (2005), Stenmark (2002), Nonaka (1991), Mulholland et al. (2000)).

8.4 Learning

Theories of learning are of course essentially linked to theories of knowledge, as learning can be seen as both knowledge development and knowledge sharing. Learning can be conceptualized in different ways. One is the essential question of where learning takes place. There are three types of learning: reflection-in-action (at the individual level); domain construction and community of practice learning (at the group level) and perspective taking (at the organizational level) (Mulholland et al. (2000)).

For example, an individual within a project team may initially create a new idea. Through collaboration within the project team, this idea may be modified and further prepared and become incorporated into revised work practices. This may become known and adopted by other project teams or parts of the organization and eventually lead to changes in company policy. At each of the steps in this example the knowledge that started as a new idea takes different forms, plays different roles, and requires different kinds of support.

One well-developed theoretical basis for understanding individual learning in a context of work, is Schön's (1983) theory of 'reflection in action'. According to Schön a professional worker (i.e. knowledge intensive worker), seemingly effortless, uses and applies a range of knowledge and skills during their professional duties. This is done in a way that could confuse an untrained onlooker. Schön refers to this as 'knowledge in action'. Reflection in action occurs when the knowledge that a worker uses and applies during their professional duties produces an unexpected outcome. If their knowledge has fallen short of dealing with the problem they were trying to solve, they may need to view the problem from a different perspective. This allows them to reflect on their actions, question their assumptions and reshape their activity, whilst in the process of their everyday work. Once the breakdown has been resolved, they can return to working with knowledge in action, combined with explicit knowledge. Reflection in action highlights an important distinction between tacit and explicit knowledge applied within the organization. Reflection in action can be thought of as a process of generating and modifying explicit knowledge related to action that was guided by tacit knowledge. Explicit knowledge has the advantage that it is open to scrutiny and can be discussed among colleagues. Individual learning and knowledge creation is therefore closely related to the interplay between tacit and explicit knowledge (Mulholland et al. (2000)).

Group level learning within organizations could be interpreted by using the notions of domain construction (Sumner (1995)) and community of practice learning (Brown and Duguid (1991)). The term 'group' is intended to encompass both officially sanctioned groups within the organization, often referred to as teams, and also unofficial groups emerging around shared interests, usually referred to as communities of practice.

Domain construction describes the processes by which groups elaborate and evolve a shared knowledge of their domain over time. Domain construction takes many forms: negotiating and defining new domain concepts, making links between concepts, and making links between concepts and the tools of work. These new concepts will evolve over time. Initially, a new concept will start to be used in informal dialogue between workers. During these dialogues, debates will occur as to the utility of the concept, its precise meaning, and its relation to other domain concepts. If the concept is found to be useful, then a consensus will be reached, and the concept becomes a more fixed and formal part of the domain. During domain construction, concepts gradually evolve from being informal and ill defined, to being formal and well defined. Essentially, domain construction describes

the process by which a group explicates and formalizes shared knowledge (Mulholland et al. (2000)).

'Community of practice learning' describes learning processes where both the existence of the community as well as the interaction between members is based on shared understanding of practices.

Organizational learning focuses on the way people make sense of their experiences at work. These experiences may derive from explicit sources, or they may be derived from tacit sources. From this view, learning is something that can emerge from social interactions, normally in the natural work setting. In the case of explicit information it involves a joint process of making sense of data. Tacit and embodied forms of learning involve situated practices, observation and emulation of skilled practitioners and socialization into a community of practice.

At the same time as we may make extinctions to identify where learning take place, we may also make distinction between learning at different levels of change. The first kind of learning is obtaining know-how in order to solve specific problems based upon existing premises. This is often referred to as Single Loop learning (Argyris and Schön (1996), Nonaka and Takeuchi (1995)). The second kind of learning is establishing new premises (i.e. mental models or perspectives) to override the existing ones. This is often referred to as Double Loop learning (Argyris and Schön (1996), Nonaka and Takeuchi (1995)).

8.5 Organizational challenges

As pointed to in the introduction, knowledge management in most large companies involve typical challenges:

- One is to be able to reduce the vulnerability with respect to loss of knowledge when people change position or leave the organization. As the previous discussion indicates, the nature of this challenge becomes even more apparent if we recognize the important knowledge as essentially tacit and thus difficult to simply store in information/knowledge databases.

- The second challenge is sharing of knowledge. This implies both the willingness to share, if and when people find their personal expertise as important for individual incentives (Wilson (2002)). At the same time, which is more widely recognized, knowledge that is tacit and/or domain specific might by its very nature be difficult to share with

others. In these settings effective tools to mediate knowledge sharing in and between practices and domains is crucial.

- The third challenge is to be able to 'translate' knowledge between the complex nature often understood within a certain (i.e. technological) discipline, into a more simple but yet relevant understanding which can guide effective practice for people outside the discipline but within a relevant practice.

In the next section, we will explore how insights into the practices of models and simulation can be used to develop more effective means to meet the challenges outlined above.

8.6 Modeling and simulation

While knowledge has been a topic of previous chapters, it is appropriate to define what modeling and simulation mean. Modeling and simulation involves two different parts, the model and the simulation.

A model (abstract or conceptual) is a theoretical representation of a physical, biological or social process or system. Traditionally, the formal modeling of systems has been via a mathematical model, which attempts to find analytical solutions to problems that enable the prediction of the behavior of the system from a set of parameters and initial conditions. Usually the model is based on the mathematical language to describe the behavior of a system with a set of variables and a set of logical and quantitative relationships between them. Models in this sense are constructed to enable reasoning within an idealized logical framework about these processes. Idealized here means that the model may make assumptions and simplifications that are known to be false in some detail, but still produce acceptably accurate solutions. Models are used primarily as a reusable tool for discovering new facts, providing systematic logical arguments, as explicatory or pedagogical aids and for evaluating hypotheses theoretically or devising experimental procedures to test them. Reasoning within models is determined by a set of logical principles, although rarely is the reasoning used completely mathematical (LaborLawTalk (2005)).

A simulation is an imitation of some real device or state of affairs. Simulation attempts to represent certain features of the behavior of a physical or abstract system by the behavior of another system. Simulation is used in many contexts, including the modeling of natural systems, and human

systems to gain insight into the operation of those systems or to test some real-world practical scenario (LaborLawTalk (2005)).

8.7 Modeling and simulation in the context of learning, knowledge creating and knowledge sharing

The word science is derived from the Latin word scientia, which means knowledge. Mathematics is essential to science, and since mathematics is fundamental to modeling and simulation, we need to understand what is meant by knowledge in order to understand what kind of knowledge is used in modeling and simulation.

In the process industry processes can be expressed e.g. by using chemistry and physics (first principle) and/or grey/black box modeling. Also operation practices can be important to model. It should be evident that at least explicit knowledge is knowledge used in developing a mathematical model. Since explicit knowledge is codifiable, it should be suitable for using in computers and hence in simulation. The question is if, and in that case how, tacit knowledge can be conserved by using modeling and simulation.

If we accept Nonaka & Takeuchi's expansion of Polanyi's 'tacit knowing' concept ((Nonaka and Takeuchi, 1995, p.60)), that is, tacit knowledge can be converted to explicit knowledge; it is most likely that it should be possible to put the knowledge into a model. Following the idea of the objectivists' camp on knowledge, the model should be easy to use and everybody should be able to understand it once finished. In practice this is not the case.

However, if we think of tacit knowledge as knowledge that is not possible to articulate, there are some challenges.

The way this is usually done in practice is that one or several employees are experts on the particular model. There are several reasons for this. One is that once articulated and coded, others do not easily understand the model if not proper trained. Another issue is that the expert over time will acquire tacit knowledge, and will have the possibility to act as a master in transferring both explicit and tacit knowledge to other employees in the organization, given a learning situation.

8.7.1 Knowledge sharing

'If getting promotion, or holding your job, or finding a new one is based on the knowledge you possess - what incentive is there to reveal that knowledge and share it?' (Wilson (2002)). As Wilson indicates, it could not be taken for granted that all employees would like to share their knowledge.

The use of modeling and simulation is traditionally seen as a tool to solve complex problems, and that the development of mathematical models for that purpose is considered expensive. Usually the development of mathematical models requires knowledge from several different disciplines. Especially in the process industry where some employees would typically be experts within physics, some in chemistry and others in modeling and simulation. Because of the diversity in knowledge, they may need to cooperate in order to develop and review the content of a mathematical model to be used in simulations. By stating that mathematical models are going to be developed and used in a problem solving process, the management have put a strong incentive and guidance to the employees to contribute to the knowledge put into and extracted from the model. This could be seen as an important and powerful strategic organizational decision in order to motivate and force the employees to share knowledge. By developing a model for simulation, the employees have to share knowledge in order to produce a usable model. The strategic decision of developing mathematical models to be used for modeling and simulation therefore contribute to sharing of knowledge (and thereby learning), to de-personalize knowledge and thereby influence the possibility for employees to hoarding their knowledge and use it in a 'knowledge is power' setting.

In addition to the benefit of knowledge sharing, the organization will not be that vulnerable of employees leaving the organization, because, at least their explicit knowledge is kept in the model – or conserved. Also if the development phase of the model has been successful, it is most likely that tacit knowledge as well has been shared due to socialization processes (Nonaka and Takeuchi (1995)) or in the domain construction process (Sumner (1995)) – which make the total knowledge even more de-personalized.

Another interesting role of modeling and simulation is the possibilities that open up when used in learning and knowledge creation. This could be illustrated by a case from practice[1].

[1] The case descriptions for Case I and II are based on experiences from the work with developing the control structure and knowledge described in (Kolås and Støre 2008) and (Kolås 2007). The period of the work process covered was from 1998 till 2001.

Case I: Modeling, Simulation, Learning and knowledge creation
The goal of a project was to reduce process variations in an electrochemical process. The company has plants spread around the globe, and each plant is organized the traditional way with operators, supervisors and local technical personell, together with one R&D department servicing all the plants. The electrochemical process is considered very complex, hard to understand, and demands continuos operation 24:7.

The official group, the project team, was composed of people from R&D, from the local organization of operation and local technicians. Once a week the project team had formal meetings, discussing various challenges and results concerning the project.

The R&D members were located almost next door to one another. Some R&D members of the project team had experience from developing and using various mathematical models of the process. During this work practice they had gained expert knowledge of the process. They all had different knowledge and background, but with some knowledge in common.

The operational 'community of practice' had built into them practice based knowledge and 'best practice' from the rest of the company.

By heavily question established 'facts' and challenging each other's knowledge and arguments, the project team developed new hypothesis that in turn was modeled and simulated in a dynamic (mathematical) model environment. Based on the results from the simulations some hypothesis was rejected and some accepted, and others emerged. The acceptance/rejection discussion was based on the simulation results which where discussed based on each project member's explicit and tacit knowledge.

The result of this work practice was not only that knowledge was shared, but also that new knowledge was created (see Kolås (2007a)) and a new way of controlling the process emerged (see Kolås and Støre (2008)).

In 'Case I' it is evident that a 'community of practice' (Argyris and Schön (1996)) emerged between some of the members of the project team. Nonaka & Takeuchi (1995) point to that an important characterization of creating knowledge is the ability of the organization to question the establishment. The R&D people were located almost next door to one another, and it was easy to have informal meetings discussing various questions and solutions. The discussions in the R&D 'community of practice' often took the form of heavily challenging established 'facts' and each other's 'knowledge', or as Senge (1990) put it – use the tension to something creative. Based on these discussions the participants were able to learn from each other and give birth to new knowledge. To test the knowledge created it

was modeled and simulated, and the simulation results acted as a basis for further discussions. The participants from the organization of operations could be seen as another 'community of practice', because of the nature of their location and collaboration practice. This 'community of practice' had built into them practice based knowledge 'collected' within the last 40 years, and 'best practice' from the rest of the company.

Kim and Senge (1994) point to that modeling and simulation could be used to 'foster shared understanding of dynamically complex policy and strategy issues', where the management used simulations to investigate the effect of various decisions in order to make the most correct one. Although knowledge creation is not the focus of Kim & Senge, the example illustrate that modeling and simulation could be used in the 'knowledge spiral' process described as 'socialization' (Nonaka and Takeuchi (1995)), where one of the purposes is to develop a shared mental model.

As in the case of Kim & Senge one of the roles of the use of modeling and simulation in 'Case I' has been to contribute to the socialization process (or the domain construction process as described in the previous section regard group level learning) to build a shared mental model among both the R&D 'community of practice' and all the members of the project team. Further, by using the model and playing with the various hypotheses, a concept of how to solve the problem emerged (externalizing). By implementing it in the model the concept could be simulated (combination) to see how the process reacted. By simulation, the project team got a kind of operational knowledge, and the knowledge created was internalized in the project team.

Another important role of the modeling and simulation is that it could be looked at as a kind of standardization (Star (1994)), in that it standardizes how to interpret the behavior of the electrochemical process. On the one side it acts as a de-personalizing of the knowledge so that it could be seen as objective knowledge. On the other side it is an instrument of power, shaping others opinion of how to interpret, in this case, the electrochemical process.

8.7.2 Simplification of knowledge

Developing mathematical models for simulation often involves the use of Partial Differential Equations (PDE's). This is also the case when modeling an electrochemical process. It is not the intention that every employee should be an expert in solving PDE's or experts in physics or chemistry, but it is expected that the operators of the process, being skilled workers, should have a basic understanding of the knowledge expressed in the model.

The necessity of simplification of knowledge has been pointed to by e.g. Zuboff (1988); '... The know-how of managers with years of experience could be systematized and made available to operators who would never have the same degree of involvement in the action contexts that had developed the personal and specific knowledge associated with action-centered skill'.

In the example of Kim and Senge (1994) referred to above, the modeling and simulation served the purpose of simplification of complex policy and strategy issues to an understandable picture.

The same was the case in 'Case I' where the output from the simulation was used to investigate and understand the behavior of the process not otherwise obvious due to complexity.

Further Zuboff (1988) point to a similar approach where '... Operators needed the kind of understanding that would allow them to know when and how to use a model, and when to be critical of its assumptions or outputs'.

In 'Case I', since all the members in the project group did not have an academic background, the simplification of knowledge served as a way of making complex information understandable so that one where able to discuss with people with a more practice-based background than academic background. Based on these discussions, the new knowledge was internalized in the project team in a democratic way.

Another important matter concerning the complexity of the electrochemical process is the time scale. If an operator initiates a change in the control inputs of the process, the dynamics of the process is such that there will go fourteen days before the changes have reached steady state. This complicates the learning and knowledge creation process, since it is very difficult to overview the effect of the changes because of both the long time elapsed, and the number of different people that in the mean time has been involved. This is pointed to by Kim and Senge (1994) which state that; '... Very often, recognizing dynamic complexity demands changes in prevailing mental models. Few organizations, in our experience have the capability to build shared understanding of dynamic complexity, ...'. This is an important side of modeling and simulation in that it is 'easy' to get the overview of how the process react to different actions, both short term and long term by doing simulations. Done right, it could contribute to building a better, shared mental model within the organization.

As stated above, the new knowledge was internalized in the project team, but when internalizing the knowledge in large scale there was some resistance in the organization. This could be illustrated with an example:

Case II: Modeling and simulation for enhancing organizational learning *When commissioning the knowledge gained in Case I, the project team experienced some resistance in the organization of operation accepting the new knowledge. The project team had developed their knowledge by working on a small amount of process units (10%), while the commissioning included the complete production line.*

The new knowledge was very different from earlier practice, and did slightly reduce the operator's freedom to manually manipulate the process. Even though all persons involved had been informed during the development and trial phase there were still some skepticism.

Not until a major part of the operators had been through a detailed learning process, learning about the physics and chemistry of the process, the new knowledge was accepted. One of the operators commented, when realizing what the new knowledge was all about: '... but then we have done it wrong for 40 years!'

We believe there are several reasons why there was some resistance. The most important one is that the new knowledge was completely different from what was common practice.

Working with modeling and simulation in order to understand the complexity of the process had resulted in new knowledge of the process behavior. Even though there were used outputs from the simulation (simplified knowledge) and results from practical tests showing that the new knowledge was working, it was hard for the operators to believe and accept it. The project team's shared model was completely different to the process model shared by the operators.

Another important part of the reason why the project experienced some resistance in the organization of operations is believed to be in that the new knowledge slightly reduced the operator's freedom to make manual adjustments to the process. This violated the idea of the 'Scandinavian democracy' and the principle of participation (Ehn (1992)). As stated in Ehn (1992) concerning an agreement on the introduction and commissioning of new computer based system in an organization; '... The orientation must be given in a well-arranged form and in a language that can be understood by non-specialists'. Since the simplified knowledge had served very well in the creation and sharing of knowledge within the project team, it was assumed that the same technique could be used when internalizing the knowledge in a large scale. But the project team did not succeed in communicating the new knowledge at this stage. What was understood by the organization of operation was that the new knowledge restricted their freedom of choice.

When talking about mental models Senge (1990) point to this as '... new insight fail to get put into practice because they conflict with deeply held internal images of how the world works, images that limit us to familiar ways of thinking and acting'. Because of this the project team temporarily had to remove this restriction from the implemented solution.

The project team realized that there was a lack of knowledge within the organization of operation, and pushed onto the management of operations in order to 'upgrade' the knowledge within the organization of operation. It was not until a major part of the operators had been through an extensive socialized learning of the physics and chemistry of the process that the new theory was accepted (internalized) in the organization. They were then able to understand the new knowledge communicated.

It is also important to mention that the training was done by some external to the project, in that it was believed that it was judiciously to have some seen as neutral to the project explaining what the new knowledge was all about.

In the practice-based camp of both knowledge and learning, knowledge creating and learning is seen as something evolved and developed by humans acting together (Nonaka and Takeuchi (1995), Argyris and Schön (1996), Hislop (2005)). Hislop (2005) point to that this necessarily is not without problems. As mentioned in 'Case II', one of the operators realized that the operational practice had been 'wrong for 40 years', so even if the knowledge is practise-based there is no guarantee that it is the right knowledge.

Nonaka and Takeuchi (1995) claims that one of the reasons to the success of Japanese companies is that they managed to adopt and make benefit of new external knowledge. Hislop (2005) points out that for an organization to learn and create knowledge it is important not to become 'inward-looking and unreceptive to ideas generated outside the 'community". If this is the case one will most likely go into a 'knowledge trap'. There is a possibility that the reason why it took a decent amount of time to gain new knowledge is that the organization was trapped in a 'knowledge trap'.

Using Argyris and Schön (1996) the situation could be described as that the organization had been trapped in the Single Loop, in that they tried to achieve the goals, but did not manage to question if they very doing the right things (the Double Loop) to achieve the goal. The reason why the organization was trapped could maybe be explained by that the operational 'community of practice' was too inward looking. Another likely part of the explanation could be that since this is an industry with an almost continuous ongoing reorganization and downsizing of the numbers

of employees, the organization of operations did simply not have the time and energy to 'reflect in action' and adopt new knowledge.

In this case, the use of modeling and simulation could be said to have been important to identify holes in the knowledge and to break a knowledge trap within the organization of operations.

8.8 Conclusion

Within the framework of this chapter we claim that using modeling and simulation conserve knowledge. By conserving knowledge an organization has the power to keep the explicit knowledge and, to some extent, parts of the tacit knowledge within the organization, making a knowledge dependent organization less vulnerable of people leaving their positions in a way or another with respect to loss of knowledge.

Conserving knowledge implies knowledge sharing, and by emphasizing modeling and simulation an organization have the opportunity to share knowledge within an organization effectively.

Also, where the knowledge is complex, the use of modelling and simulation serve as a simplification of complexity, suitable for other parts of the organization that is only interested in the behavior and not the detailed chemistry/physics. In a situation like this, the simplified knowledge could act as a source for learning and knowledge creation.

8.9 Acknowledgements

We acknowledge Professor Morten Levin, and Professor Eric Monteiro at the Norwegian University of Technology for the inspiration, discussions and feedback during the writing of this chapter. The financial support of Hydro Aluminium AS is also gratefully acknowledged.

Chapter 9

Conclusion

Aluminium is a metal playing an important role in all modern life of today, and will continue to play an important role in many years to come. As with every business and everyone who participates in the global economy, climate change is a challenge also shared by the aluminium industry. In order for the aluminium industry to meet their goals in reducing greenhouse gas emissions and increase energy efficiency in aluminium production, not only new technology and knowledge will play an important role, but also the speed in which the organization is able to utilize new technology and knowledge is crucial. E.g. an important factor for succeeding in reducing greenhouse gas emissions and increase energy efficiency, is the use of new advanced process control systems and increased process knowledge.

On the way towards new advanced state of the art process control systems, an extention of the current control paradigm may be preferred. In Chapter 2, the so-called correlation line, which is very well known in the community of aluminium production, is investigated and discussed, and it is conlcuded that, given constant alumina concentration, the correlation line is unique (otherwise not) (Kolås (2007a)). Based on the correlation line, a control structure is proposed in Chapter 3, aiming for constant mass of aluminium fluoride in the Hall- Heroult cell (Kolås and Støre (2008)). It is shown that the control structure reduces process variations and energy consumption without loss of production.

New advanced process control systems implies utilizing state of the art process control systems as e.g. Nonlinear Model Predictive Control (NMPC). In this thesis it is claimed that there are two real challenges regarding NMPC applications: 1) the state estimation and 2) the organization.

- The state estimation - because by not being able to measure all the states, the quality of the estimates determine to a large extent the performance of the NMPC application (Chapter 4, 5 and 6).

- The organization - because they are the real user of the NMPC application, and the degree of success is dependent on their ability to make use of it (Chapter 8).

Several authors have experienced shortcomings applying the EKF to systems with severe nonlinearities and/or constraints. Chapter 5 compares EKF and a selection of UKF algorithms on nonlinear process systems with multimodal probability density functions, and it is shown that the performance of the constrained UKF approach, with an algebraic reformulation of the correction part, shows very strong performance on our selection of nonlinear constrained process systems.

In Chapter 6 alternative concepts of noise modeling are investigated, and it is shown that modeling noise by introducing it in the system auxiliary variables and control inputs may have a positive impact on estimation performance.

In Chapter 7 an algorithm for computing the dynamic eigenvalues for linear time-varying (LTV) systems is proposed. The algorithm can be used to analyze the stability for some types of LTV systems.

Those responsible for commissioning an NMPC application in an organization, e.g. in the process industry, will meet an organization that at the current moment most likely do not have the knowledge to understand what NMPC is all about, which is resistant to changes and might not share the assumptions on which the model are founded. Further, they might not agree in the control actions the NMPC might take, because it does not line up with the knowledge embodied in the organization. These aspects makes the organization an important challenge, since the organization is the real user of the NMPC application and the degree of success is dependent on their ability to make use of it. The Hall-Heroult process is a knowledge intensive process, and the organization in a Hall-Heroult plant may face challenges in knowledge processes. The process experienced by producing the content in the papers Kolås and Støre (2008) and Kolås (2007a) with respect to the process involved in applying modeling and estimation from an organization point of view is addressed by Chapter 8 by looking at possibilities that opens up with the use of modeling and simulation in the context of learning, knowledge creation and knowledge sharing. These issues are addressed by analyzing the role of modeling and simulation in

the context of conserving knowledge. Amongst others, it is concluded that if the knowledge is complex, the use of modeling and simulation serve as a simplification of complexity, suitable for other parts of the organization that is only interested in the behavior and not the detailed chemistry/physics. In a situation like this, the simplified knowledge could act as a source for learning and knowledge creation. This awareness may be important when commissioning new technology as e.g. an NMPC application in the Hall-Heroult organization.

Bibliography

Aalbu, J. (1986). Adaptive Control of Alumina Reduction Cells with Point Feeders. *MIC*, 7:45–56.

Adrianova, L. Y. (1995). Introduction to Linear Systems of Differential Equations. *Arner Mathematical Society*.

Akin, B., Orguner, U., and Ersak, A. (2003). State estimation of induction motor using Unscented Kalman Filter. *Proceedings of 2003 IEEE Conference on Control Applications*, 2:915–919.

Argyris, C. and Schön, D. A. (1996). Organizational Learning II. *Addison Wesley*.

Bellatoni, J. and Dodge, K. (1967). A square root formulation of the Kalman Schmidt filter. *AIAA Journal*, 5:1309–1314.

Bizup, D. F. and Brown, D. E. (2003). The over-extended Kalman filter - dont use it! *Proceedings of the Sixth International Conference of*, 1:40–46.

Briers, M., Maskell, S. R., and Wright, R. (2003). A Rao-Blackwellised unscented Kalman filter. *Proceedings of the 6th International Conference of Information Fusion*, 1:55–61.

Brown, J. and Duguid, P. (1991). Organizational Learning and communities-of-practice: towards a unified view of working, learning and innovation. *Organization Science*, 2(1):40–57.

Carlsen, A., R, R. K., and von Krogh, G. (2004). Living Knowledge. The dynamics of professional service work. *Palgrave. London*.

Chen, C. T. (1999). Linear System Theory and Design. *Third Edition*, Oxford University Press.

Chen, J. and Taylor, M. P. (2005). Control of temperature and aluminium fluoride in aluminium reduction. *Aluminium*, 7(8):678–682.

Chen, W., Lang, L., Bakshi, B. R., Goel, P. K., and Ungarla, S. (2006). Bayesian Estimation via Sequential Monte Carlo Sampling - Constrained Dynamic Systems. *Preprint submitted to Elsevier Science*.

Daum, F. E. (1986). Beyond Kalman filters: practical design of nonlinear filters. *SPIE*, 2561:252–262.

Daum, F. E. (2005). Nonlinear Filters:Beyond the Kalman filters. *IEEE AE system magazine*, 20(8):57–69.

Desclaux, P. (1987). AlF3 additions based on bath temperature measurements. *Light Metals*, pages 309–313.

Dieci, L. (1992). Numerical integration of the differential Riccati equation and some related issues. *SIAM J. Numerical Analysis*, 29(3):781–815.

Drengstig, T. (1997). On process model representation and AlF3 dynamics of aluminium electrolysis cells. *Dr.Ing. thesis, Norwegian University of Science and Technology (NTNU), 1997:94*.

Drengstig, T., Kolås, S., and Støre, T. (2002). The Impact of Varying Conductivity on the Control of Aluminum Electrolysis Cells. *Light Metals*, pages 377–382.

Drengstig, T., Ljungquist, D., and Foss, B. (1998). On the AlF3 and temperature control of an aluminumelectrolysis cell. *IEEE Transactions on Control Systems Technology*, 6:157–171.

Economist, T. (2000). In praise of Bayes. *The Economist*, September(30).

Ehn, P. (1992). Scandinavian Design: On Participation and Skill. *Usability, Turning technologies into tools, edited by Paul S.*, Oxford University Press:96–132.

Entner, P. M. (1992). Control of AlF3 concentration. *Light Metals*, pages 369–374.

Entner, P. M. (1993). Further development of the AlF3-model. *Light Metals*, pages 265–268.

Entner, P. M. (1995). Control of Bath Temperature. *Light Metals*, pages 369–374.

Entner, P. M. and Gudmundsson, G. A. (1996). Further development of the temperature model. *Light Metals*, pages 445–449.

Foss, B. and Schei, T. S. (2005). Putting Nonlinear Model Predictive Control into Use. *Workshop on Nonlinear Model Predictive Control*, Freudenstadt(Germany).

Foss, B. and Schei, T. S. (2007). Putting Nonlinear Model Predictive Control into Use. *Assessment and Future Directions of Nonlinear Model Predictive Control, Springer Verlag*, 358:407–417.

Freiling, G. (2002). A survey of nonsymmetric Riccati equations. *Linear Algebra and its Application*, pages 243–270.

Gelb, A. (1974). Applied Optimal Estimation. *The M.I.T. Press*, ISBN 0-262-20027-9.

Gran, E. (1980). A Multi-Variable Control in Aluminum Reduction Cells. *Modeling Identification and Control*, 1(4):247–258.

Grjotheim, K. and Kvande, H. (1993). Introduction to aluminium electrolysis. *Aluminum-Verlag*.

Hao, Y., Xiong, Z., Sun, F., and Wang, X. (2007). Comparison of Unscented Kalman Filters. *Proceedings of the 2007 IEEE ICMA*, Univeristy of Wisconsin-Madison(Aug 2007):895–899.

Hasseltine, E. L. and Rawlings, J. B. (2003). A Critical evaluation of Extended Kalman Filtering and Moving Hoizon Estimation. *TWMCC, Technial report number 2002-03*.

Haupin, W. (1987). Cell Voltage. *In: The 6th International Course on Process Metallurgy of Aluminium, Institute of Inorganic Chemistry, NTH, Trondheim, Norway*, Sec. 7:1–25.

Haupin, W. and Kvande, H. (1993). Mathematical model of fluoride evolution from Hall Heroult cells. *Light Metals*, pages 257–263.

Henriksen, R. and Balchen, J. G. (1990). Multivariable stokastiske systemer. *Report 90-92-X*, NTNU(Department of Engineering Cybernetics).

Hislop, D. (2005). Knowledge management in organizations. A critical introduction. *Oxford University Press*.

Hives, J., Thonstad, J., Sterten, Å., and Fellner, F. P. (1993). Electrical conductivity of molten cryolite–based mixtures obtained with a tube-type cell made of pyrolitic boron nitride. *Light Metals*, pages 247–256.

Hyland, M. M., Patterson, E. C., Stevens-McFadden, F., and Welch, B. J. (2001). Aluminium fluoride consumption and control in smelting cells. *Scandinavian Journal Of Metallurgy*, 30:404–414.

IAI (2008). International Aluminium Institute. *http://www.world-aluminium.org/ as pr. 11. april 2008*.

Julier, J. and Laviola, J. J. (2007). On Kalman Filtering With Nonlinear Equality Constraints. *IEEE Transactions on Signal Processing*, 55(6):2774–2784.

Julier, S. J. (2002). The Scaled Unscented Transform. *Proceedings of the American Control Conference*, 6:4555–4559.

Julier, S. J. (2003). The spherical simplex unscented transformation. *Proceedings of the American Control Conference, June 4-6*, pages 2430–2434.

Julier, S. J. and Uhlmann, J. K. (1994). A general method for approximating nonlinear transformation of probability distributions. *Technical report, RRG, Dept. of Engineering Science, University of Oxford*, Can be downloaded from http://www.robots.ox.ac.uk/˜siju/work/publications/letter size/Unscented.zip, August 1994.

Julier, S. J. and Uhlmann, J. K. (2002). Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. *Proceedings of the American Control Conference, May 8-10*, pages 887–892.

Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings IEEE, March 2004*, 92(3):401–422.

Julier, S. J., Uhlmann, J. K., and Durrant-Whyte, H. (1995). A new approach for filtering nonlinear system. *Proceedings of the American Control Conference*, pages 1628–1632.

Julier, S. J., Uhlmann, J. K., and Durrant-Whyte, H. (1997). A new extension of the Kalman filter to nonlinear systems. *Proceedings of Aerospace/Sense: The 11th Int. Symp. On Aerospace/Defence Sensing, Simulation and Controls*, pages 1628–1632.

Julier, S. J., Uhlmann, J. K., and Durrant-Whyte, H. (2000). A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. on Automatic Control*, 45(3):477–482.

Kalman, R. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transaction of the ASME-Journal of Basic Engineering*, 82(Series D):35–45.

Kamen, E. W. (1988). The poles and zeros of a linear time-varying system. *Lin. Algeb. Appl.*, 98:263–289.

Kandepu, R., Foss, B., and Imsland, L. (2008). Applying the Unscented Kalman Filter for Nonlinear State Estimation. *Journal of Process Control*, 18:753–768.

Khalil, H. K. (1996). Nonlinear Systems. *Prentice Hall,*, Third Edition.

Kim, D. and Senge, P. M. (1994). Putting systems thinking into practice. *System Dynamics Review*, 10(2-3):277–290.

Klev, R. and Carlsen, A. (2000). Meninger og myter om kunnskap. I M.Rolfsen (red): Trendenes tyranni. Produksjon og arbeid i et nytt århundre. *Bergen.*, Fagbokforlaget.

Kolås, S. (2007a). Defining and Veryfying the 'Correlation Line' in Aluminium Electrolysis. *JOM Journal of the Minerals, Metals and Materials Society*, 59(5):55–60.

Kolås, S. (2007b). Method and means for controlling an electrolysis cell. *Norsk Hydro Patent Application*, NO 20075933.

Kolås, S., Foss, B., and Schei, T. S. (2008a). Constrained nonlinear state estimation based on the UKF approach. *to Computers and Chemichal Engineering*, Submitted 2008.

Kolås, S., Foss, B., and Schei, T. S. (2008b). Noise modeling concepts in Nonlinear State Estimation. *Journal of Process Control*, Submitted 2008.

Kolås, S., Foss, B. A., and Schei, T. S. (2008c). State estimation IS the real challenge in NMPC. *to Proceedings to NMPC 08, Pavia, Italy.*

Kolås, S., Neerhoff, F. L., van der Kloet, P., and Foss, B. (2008d). A recursive algorithm for computing the dynamic eigenvalues. *Submitted to Systems & Control Letters.*

Kolås, S. and Støre, T. (2008). Bath Temperature and AlF3 Control of an Aluminium Electrolysis Cell. *Submitted to Control Engineering Practice*.

LaborLawTalk (2005). http://encyclopedia.laborlawtalk.com. *as pr. 2005.04.12*.

Laviola, J. J. (2003). A comparison of Unscented and Extended Kalman filtering for estimating quaternion motion. *American Control Conference, 2003. Proceedings*, 3:2435–2440.

Li, W. and Leung, H. (2004). Simultaneous Registration and Fusio of Multiple Dissimilar Sensors for Cooperative Driving. *IEEE Transactions on Intelligent Transportations Systems*, 5(2):84–98.

Ljung, L. (1987). System Identification. Theory for the User. *Prentice-Hall*, Englewodd Cliffs(NJ).

Marafioti, G., Hovd, M., and Olaru, S. (2008). State Estimation in Nonlinear Model Predictive Control, Unscented Kalman Filter Advantages. *NMPC 08, Pavia, Italy*.

Meghlaoui, A., Al Farsi, Y. A., and Aljabri, N. H. (2002). Analytical and Experimental Study of Fluoride Evolution. *Light Metals*.

Meghlaoui, A., Thibault, J., Bui, R. T., Tikasz, L., and Santerre, R. (1998). Predictive control of aluminum electrolytic cells using neural networks. *Metallurgical and Materials Transactions*, 10(1).

Mulholland, P., Domingue, J., and Hatala, Z. Z. M. (2000). Organisational Learning: An Overview of the Enrich Approach. *Journal of Information Services and Use*, 20(1):9–23.

Neerhoff, F. L. and van der Kloet, P. (2000). Frequency behavior of time-varying small-signal models of nonlinear circuits. *Proc. NDES 2000*, Catania, Italy(May 18-20):81–84.

Neerhoff, F. L. and van der Kloet, P. (2001). A complementary view on time-varying systems. *IEEE International Symposium on Circuits and Systems, Sydney, Australia*, III:779–782.

Nonaka, I. (1991). The knoweledge-creating company. *Harvard Business Review, Nov-Dec*.

Nonaka, I. and Takeuchi, H. (1995). The knoweledge-creating company. *Oxford University Press*.

Nørgaard, M., Poulsen, N. K., and Ravn, O. (2000). New developments in state estimation for nonlinear systems. *Automatica*, 36:1627–1638.

Pannoncchia, G. (2003). Robust disturbance modeling for model predictive control with application to multivariable ill-conditioned processes. *Journal of Process Control*, 13(8):693–701.

Peyneau, J. (1988). The automated control of bath composition on high amperage cell. *In: Proceedings of International Symposium on Reduction and Casting of Aluminium*, pages 189–195.

Pieper, R. J. B. (2007). Comparing Estimation Algorithms for Camera Position and Orientation. *Department of Electrical Engineering,Linköpings universitet, SE581 83 Linköping, Sweden*.

Polanyi, M. (1983). Tacit dimension. *Gloucester, Mass. Peter Smith 1983 (reprinted), First published 1966*.

Rao, C. V. (2000). Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete-Time Systems. *PhD thesis (Chemical Engineering)*, Univeristy of Wisconsin-Madison.

Rawlings, J. B. and Bakshi, B. R. (2006). Particle filter and moving horizon estimation. *Computers and Chemichal Engineering*, 30:1529–1541.

Rawlings, J. B., Jerome, N. F., Hamer, J. W., and Bruemmer, T. M. (1989). End-point control in semi-batch chemical reactors. *Proceedings of the IFAC Symposium on Dynamics and Control of Chemichal Reactors, Distillation Columns and Batch Processes*, pages 323–328.

Rieck, T., Iffert, M., White, P., Rodrigo, R., and Kelchtermans, R. (2003). Increased Current Efficiency and Reduced Energy Consumption at the TRIMET Smelter Essen using 9 Box Matrix Control. *Light Metals*, pages 449–456.

Romanenko, A. and Castro, J. A. A. M. (2004). The unscented filter as an alternative to the EKF for nolinear state estimation: A simulation case study. *Computers and Chemical Engineering*, 28:347–355.

Romanenko, A., Santos, L. O., and Afonso, P. A. F. N. A. (2004). Unscented kalman filtering of a simulated pH system. *Ind. Eng. Chem. Res*, 43:7531–7538.

Saksvikrønning, T., Gran, E., and Vee, K. (1976). Estimation of States in Aluminum Reduction Cells Applying Extended Kalman Filtering Algorithm Together with a Nonlinear Dynamic Model and Direct Measurements. *TMS*, AIME-meeting, Las Vegas, USA.

Salt, D. J. (1990). Bath Chemistry Control System. *Light Metals*, pages 299–304.

Schei, T. S. (1997). A Finite-Difference Method for Linearization in Nonlinear Algorithms. *Automatica*, 33(11):2053–2058.

Schiff, J. and Schnider, A. (1999). A natural approach to the numerical integration of Riccati differential equations. *SIAM Journal Of Numerical Analysis*, 36(5):1392–1413.

Schuler, H. and Schmidt, C. U. (1992). Calorimetric state estimators for chemical reactor diagnosis and control: review of methods and applications. *Chem. Eng. Sci.*, 47(4):899–915.

Senge, P. M. (1990). The Fifth Discipline The Art & Practice of The Learning Organization. *Century Business*.

Shuiping, Z., Jinhong, L., and Qiangqiang, P. (2008). Model Predictive Control of Superheat for Prebake Aluminum Production Cells. *Light Metals*, pages 347–351.

Simon, D. (2006). *Optimal State Estimation*. Wiley, ISBN-10 0-471-70858-5.

Simon, D. and Simon, D. L. (2005). Aircraft Turbofan Engine Health Estimation Using Constrained Kalman Filtering. *ASME Journal of Engineering for Gas Turbine and Power*, 127:323–328.

Simon, D. and Simon, D. L. (2006). Kalman Filtering with Inequality Constraints for Turbofan Engine Health Estimation. *IEEE Proceedings Control Theory and Applications*, 153(3):371–378.

Solheim, A. (2005). Personal communication. *Trondheim, June*.

Solheim, A. (2008). Personal communication. *Trondheim, June*.

Solheim, A. and Støen, L. (2005). Personal communication. *NTNU, SINTEF*.

Spender, J. C. (1993). Competitive advantage from tacit knowledge? Unpacking the concept and its strategic implication. *Academy of Management Best Paper Proceedings*, pages 37–41.

Star, G. B. S. L. (1994). Knowledge and infrastructure in international information management: Problems of classification and coding. *In L. Bud-Frierman (ed), Information acumen: The understanding and use of knowledge in modern business*, London:(Routledge):187–216.

Stenmark, D. (2002). Information vs. Knowledge: The Role of intranets in Knowledge Management. *In Proceedings of HICSS-35, IEEE Press, Hawaii, January 7-10*.

Stevens Mc Fadden, F. J., Bearne, G. P., Austin, P. C., and Welch, B. J. (2001). Application of advanced process control to aluminium reduction cells–a review. *Light Metals*, pages 1233–1242.

Stevens Mc Fadden, F. J., Welch, B. J., and Austin, P. C. (2006). Multivariable Model-Based Control of the Non-Alumina Electrolyte Variables in Aluminum Smelting Cells. *Journal Of Metal*, 58(2):42–47.

Sumner, T. (1995). The high-tech toolbelt: A study of designers in the workplace. *Human Factors in Computing Systems (CHI'95), May 7-11, Denver, CO*.

Søderstrøm, T. and Stoica, P. (1989). System identification. *Prentice Hall*.

Taylor, M. P. (1992). Fluoride material balance. *In: Fourth Australian Aluminium Smelter Technology Workshop, Sydney, Australia*, pages 720–732.

Tenne, D. and Singh, T. (2003). The higher order unscented filter. *Proceedings of the American Control Conference*, 3:2441–2446.

Thonstad, J. and Rolseth, S. (1983). Equilibrium between bath and side ledge. *Light Metals*, pages 415–424.

Vachhani, P., Narasimhan, S., and Rengaswamy, R. (2006). Robust and reliable estimation via Unscented Recursive Nonlinear Dynamic Data Reconciliation. *Jounal of Process Control*, 16:1075–1086.

Vachhani, P., Rengaswamy, R., Gangwal, V., and Narasimhan, S. (2005). Recursive Estimation in Constrained Nonlinear Dynamical Systems. *AIChE Journal*, 51(3):946–959.

van der Kloet, P. and Neerhoff, F. L. (1999). Behaviour of dynamic eigenpairs in slowly varying systems. *7th International Workshop on Nonlinear Dynamics of Electronic Systems*, Rønne, Denmark.

van der Kloet, P. and Neerhoff, F. L. (2000). Diagonalization algorithms for linear time-varying dynamic systems. *International Journal of System Science*, 31(8):1053–1057.

van der Kloet, P. and Neerhoff, F. L. (2001). Modal factorisation of time varying models for non-linear circuits by the Riccati transform. *IEEE International Symposium on Circuits and Systems 2001*, III:553–556.

van der Kloet, P. and Neerhoff, F. L. (2004). On characteristic equations, dynamic eigenvalues, Lyapunov exponent and Floquet numbers for linear time-varying systems. *Proc. MTNS*, Leuven, Belgium, July 5-9.

van der Kloet, P., Neerhoff, F. L., and Waning, N. H. (2007). Some analytical calculations of the characteristic exponents. *International Journal of Bifurcation and Chaos*, 17(10):3673–3678.

van der Merwe, R. (2004). Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models. *PhD dissertation submitted to the faculty of the OGI School of Science and Engineering at Oregon Health and Science University.*

von Krogh, G., Ichijo, K., and Nonaka, I. (2000). Enabling Knowledge Creation: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation. *Oxford University Press, New York and Oxford.*

Welch, B. J. (1999). Aluminum Production Paths in the New Millennium. *JOM*, 51(5):24–28.

Wilson, M. J. (1992). Practical considerations used in the development of a method for calculating aluminium fluoride based on cell temperature. *Light Metals*, pages 375–378.

Wilson, T. D. (2002). The nonsense of 'knowledge management'. *Information Research*, 8(1).

Wu, M. Y. (1974). A note on stability of linear time-varying systems. *IEEE Transactions on Automatic Control*, April 1974:162.

Wu, M. Y. (1980). A new concept of eigenvalues and eigenvectors and its applications. *IEEE Transactions on Automatic Control*, Vol.AC-25(No 24):824–826.

Wu, M. Y. (1984). On stability of linear time-varying systems. *Int. J. Systems Science*, 15(2):137–150.

Xiong, K., Chang, H. Y., and Chan, C. W. (2006). Performance evaluation of UKF-based nonlinear filtering. *Automatica*, 42:261–270.

Xiong, K., Chang, H. Y., and Chan, C. W. (2007). Authors reply to "Comments on 'Performance evaluation of UKF-based nonlinear filtering'". *Automatica*, 43:569–570.

Zandt, J. R. V. (2001). A more robust unscented transform. *SPIE Proceedings of AeroSense: Signal and Data Processing of Small Targets*, 4473:371–380.

Zhu, J. and Johnson, C. D. (1989). New results in the reduction of linear time-varying dynamical systems. *SIAM J. Control and Optimization*, 27(3):476–494.

Zhu, J. and Johnson, C. D. (1991). Unified canonical forms for matrices over a differential ring. *Lin. Algeb. Appl*, 147:201–248.

Zhu, J. and Johnson, C. D. (1997). Pd-spectral theory for multi-input-multioutput linear time-varying systems. *Proc. 36th Conf. Decis. and Control*, pages 3908–3913.

Zuboff, S. (1988). In the age of the smart machine, The future of work and power. *Chap 11, New York, Basic Book.*