One cannot learn anything so well as by experiencing it oneself.

– Albert Einstein

University of Alberta

Filtering Approaches for Inequality Constrained Parameter Estimation

by

Xiongtan Yang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Chemical Engineering

Department of Chemical and Materials Engineering

Spring 2013
Edmonton, Alberta

This thesis is dedicated to my parents.

Abstract

Parameter estimation of a dynamic system is an important task in process systems engineering. The utilization of an augmented system offers the approach of estimating process states and parameters simultaneously. In practice, the parameters often satisfy certain constraints which should be incorporated to improve the estimation performance. This thesis focuses on the inequality constrained parameter estimation problem. We introduce a method of constructing inequality constraints on parameters from routine steady-state operation data. A constraint implementation method with the unscented Kalman filter (UKF) is proposed that yields faster recovery of parameter estimates than the conventional projection method. The appropriate use of projection method with the ensemble Kalman filter (EnKF) is introduced. Also, a constrained estimation method with the EnKF is proposed which results in improved performance compared to the projection method. For the moving horizon estimation (MHE), we propose an alternative approach for constrained parameter estimation, which provides better performance than the directly constrained MHE. The efficacies of the proposed approaches in this thesis are evaluated using several simulated process examples.

Acknowledgements

Standing at the point of finishing all the research work, I am very grateful to have so many people who have offered their generous help in my way of pursuing my M.Sc. degree. All those days spent over the last two years will definitely become one of the most meaningful experiences in my life.

I would like to express my thanks and appreciation to Dr. Biao Huang for his outstanding supervision. His excellence insight in academics gave me valuable directions of my research. His dedication to a quality research and work provided major source of inspiration to me. I would also like to express my thanks and appreciation to Dr. Vinay Prasad for his devoted guidance. He provided in-depth discussion and suggestions for my research. His encouragement provided me the room of improvement. They always spent significant amount of time to give me a careful review and detailed correction of my writings. Their commitment to my research is the key for my fulfillment here and two years learning experience with them has become memorable experience for me.

The Computer Process Control group provide a great working environment for discussing and sharing ideas. Special thanks go to Yijia Zhu, Li Xie, Aditya Tulsyan and Yalin Wang, who provided me valuable aids and discussions on my research work.

I would like to acknowledge the Department of Chemical and Materials Engineering at University of Alberta for providing me the opportunity to pursue my master degree. Financial support from the NSERC Industrial Research Chair in Control of Oil Sands Processes is gratefully acknowledged.

Last but not least, I would like to express my undescribable gratitude and deepest love to my parents for their understanding and unconditional support. This thesis is dedicated to my beloved parents.

Contents

List of Figures

List of Tables

List of Symbols

x	state variable
$\hat{x}$	state estimate
x_a	augmented state variable of state and parameter
x^a	augmented state variable in the UKF
$f(\cdot)$	nonlinear state transition function
$h(\cdot)$	nonlinear measurement transition function
θ	system parameter
w	process noise
w_p	parameter random walk noise
w_a	augmented process noise
v	measurement noise
Q	covariance of process noise w
Q_p	covariance of parameter random walk noise w_p
Q_a	covariance of augmented noise w_a
R	covariance of measurement noise v
A, B	dynamic matrices in linear state equation
H	dynamic matrix in linear measurement equation
u	system input
k	discrete time instant k
K	Kalman gain
P	covariance of state estimate $\hat{x}$
$(\cdot)^-$	prior statistical knowledge
I	identity matrix

x_n	nominal state variable
F	Jacobian matrix $\frac{\partial f}{\partial x}$
H	Jacobian matrix $\frac{\partial h}{\partial x}$
$\mathcal{X}$	sigma points
W	weights for sigma points in the UKF, consist of W^m and W^c
W^m	weights for calculating mean in the UKF
W^c	weights for calculating covariance in the UKF
P^a	augmented covariance in the UKF
$\mathcal{Y}$	transformed sigma points through h
$\hat{y}$	measurement estimate in the UKF
P^{yy}	covariance of measurement estimate
P^{xy}	cross-covariance of the state estimate and measurement estimate
L	dimension of vector in sigma points $\mathcal{X}$
T_s	sampling time
$\gamma, \lambda, \alpha, \beta, \kappa$	parameters involved in the UKF
$(\cdot)_f$	feeding condition
$c(\cdot)$	function of parameter relationship
w^i, v^i	noise particle in the EnKF
x^i	state particle in the EnKF
y^i	measurement particle in the EnKF
Pr	projection or clipping
D	difference between two variables
$p(x\|y)$	probability of x given y
$\hat{w}$	estimated process noise in MHE
$\circ$	function composition, i.e., $h \circ f = h(f(x))$
G	$\begin{bmatrix} h(x_{t-N}) \\ h \circ f(x_{t-N}) \\ \vdots \\ h \circ f \circ \ldots \circ f(x_{t-N}) \end{bmatrix}$

$\bar{x}_0$	initial guess in MHE
$\Gamma(x_0)$	initial penalty function $\|x_0 - \bar{x}_0\|^2_{P_0^{-1}}$ in MHE
$L_k(w_k, v_k)$	$\|v_k\|^2_{R^{-1}} + \|w_k\|^2_{Q^{-1}}$ in MHE
$\mathcal{Z}_{T-N}(z)$	arrival cost at time $T - N$
$\hat{x}^{mh}$	estimated trajectory by MHE
Φ	constant arrival cost

List of Abbreviations

PID	proportional-plus-integral-plus-derivative controller
MMSE	minimum mean square error
KF	Kalman filter
EKF	extended Kalman filter
UKF	unscented Kalman filter
UT	unscented transformation
PF	particle filter
EnKF	ensemble Kalman filter
MHE	moving horizon estimation
CSTR	continuous stirred tank reactor
PMMA	Poly(methyl methacrylate)
PDF	probability density function
NLP	nonlinear programming
MPC	model predictive control

Chapter 1

Introduction

1.1 Motivation and objective

In today's process industry, there is a high demand for improvement of operating performance. A significant amount of money has been invested in modern processing equipment and control systems, with the expectation of improved productivity and profitability. Most advanced process control strategies rely on the assumption that the states of the system are explicitly available. However, some real-time process states are not always available due to various reasons, such as unavailability of some measurements, device failure, and considerable level of noise. Even when these variables are available through lab analysis, such measurements often have significant time delays and/or irregular sampling intervals. In such cases, a state estimator, or a soft sensor, is required to obtain process states from available measurements. This will not only provide improved control performance, but will also assist in process monitoring and diagnosis.

A simple PID controller, which is used to control the water level in the tank, is shown in Fig.1.1. The level control performance depends a lot on the feedback signal, in which there will be some level of noise introduced by the measurement device. State estimation can be applied on this system to reduce the noise in the level measurements. As a result, one can expect improvement of control performance.

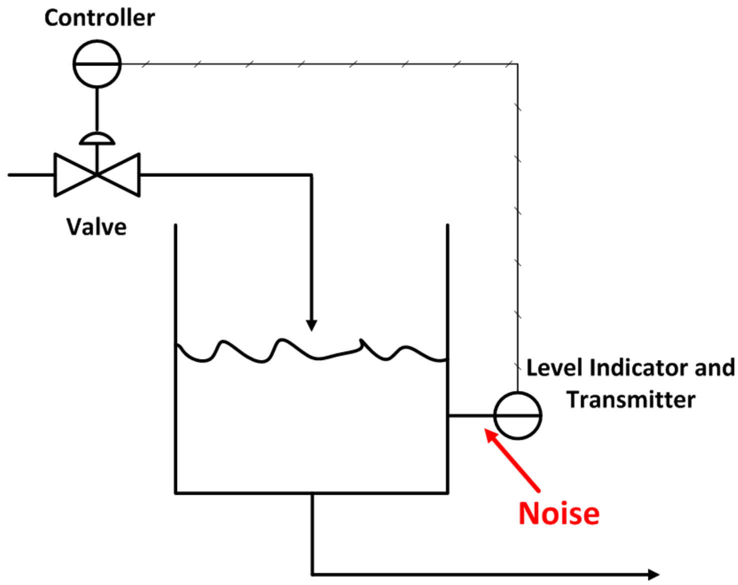


Figure 1.1: A simple level control system

In practical estimation problems, often the model structure is known but some of its parameters are unknown. Hence, we need to estimate the states and parameters at the same time with available measurements. The use of measurements to estimate parameters in a dynamic model is also an important component in the development of a predictive model for a physical process. A typical control system is shown in Fig.1.2. There is noise both in the measurement sensor and in the process model. Meanwhile, some of the parameters in the process model are unknown. The objective is to use input data u and output data y to estimate the states and parameters simultaneously. This is also called the dual estimation problem.

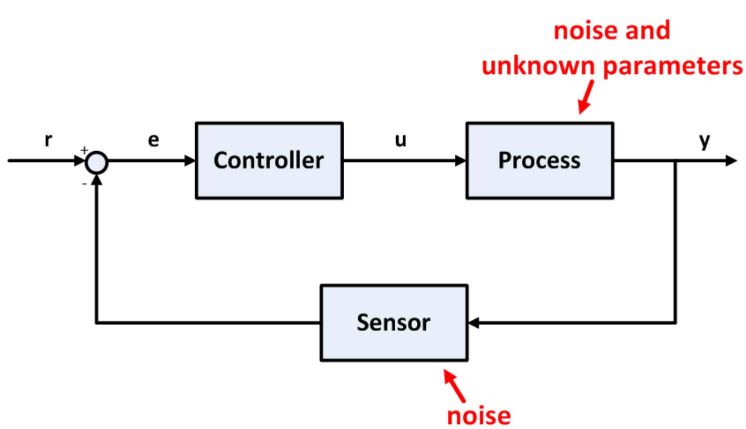


Figure 1.2: A typical control system

Most physical dynamic systems are continuous processes, while discrete measurements are taken using a measurement device with sampling time T_s. Conventionally, unmeasured disturbances have been modeled as Gaussian white noises, which are additive in both the state dynamics and measurements. Therefore, the system and measurement models are given in the following continuous-discrete stochastic form:

$$\dot{x} = f(x, u, \theta) + w \tag{1.1a}$$

$$y_k = h(x_k) + v_k \tag{1.1b}$$

where $x, x_k \in \mathbb{R}^n$ denote the vector of states, $u \in \mathbb{R}^q$ denotes the vector of manipulated variables and $y_k \in \mathbb{R}^m$ denotes the vector of available measurements. $f : \mathbb{R}^n \to \mathbb{R}^n$ is the deterministic state transition function with parameter $\theta \in \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ is the measurement transition function. f and h are not necessarily linear functions. $w \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are process and measurement noise respectively, with zero-mean independent Gaussian distributions

$$w \sim \mathcal{N}(0, Q) \tag{1.2}$$

$$v_k \sim \mathcal{N}(0, R) \tag{1.3}$$

where Q and R are covariance matrices.

In the common approach for dual estimation problem, the state vector x and parameter vector θ are combined into an augmented state $x_a \in \mathbb{R}^{n+p}$, and standard state estimation is carried out on the augmented system.

$$x_a = \begin{bmatrix} x \\ \theta \end{bmatrix} \tag{1.4}$$

Sequential parameter estimation requires a dynamic model to describe the uncertainty of the parameter θ. The standard practice is to use the dynamic parameter equation

$$\dot{\theta} = 0 + w_p \tag{1.5}$$

which is a random walk model for the parameter θ. w_p is chosen as a zero mean Gaussian noise with covariance matrix Q_p. The random walk model (Eq. 1.5) is the driving force for parameter estimation in recursive estimation algorithms. Therefore, the augmented model is

$$\dot{x} = f(x, u, \theta) + w \tag{1.6a}$$

$$\dot{\theta} = 0 + w_p \tag{1.6b}$$

$$y_k = h(x_k) + v_k \tag{1.6c}$$

Rewriting the model with augmented state x_a gives

$$\dot{x}_a = f(x_a, u) + w_a \tag{1.7a}$$

$$y_k = h(x_{ak}) + v_k \tag{1.7b}$$

where we redefine $f : \mathbb{R}^{n+p} \to \mathbb{R}^{n+p}$ as the deterministic nonlinear state transition function and $h : \mathbb{R}^{n+p} \to \mathbb{R}^m$ as the measurement transition function. $w_a = \begin{bmatrix} w \\ w_p \end{bmatrix}$ denotes the augmented state noise, which has the following distribution:

$$w_a \sim \mathcal{N}(0, Q_a) \qquad Q_a = \begin{bmatrix} Q & 0 \\ 0 & Q_p \end{bmatrix} \tag{1.8}$$

This model is a nonlinear model for augmented state x_a, because of the nonlinear combination between states and parameters. Therefore, the parameter estimation problem for either a linear process or a nonlinear process becomes a nonlinear estimation problem when augmenting the states.

Besides the available input and output data, we might have additional information about a system. For instance, some information comes from physical limitations of the process, e.g., estimated concentrations should remain positive. Also, we could have other additional information about the process of interest, such as steady-state operating data. Such additional information could be transformed into equality or inequality constraints on parameters.

An equality constraint on parameters can be written as

$$c(\theta_1, \theta_2, \ldots, \theta_p) = d \tag{1.9}$$

where p indicates the number of parameters, i.e., $\theta \in \mathbb{R}^p$. $c(\cdot)$ is a function describing the parameter relationship. In the following content, $c(\theta_1, \theta_2, \ldots, \theta_p)$ is written as $c(\theta)$ for notation simplicity. Normally, we cannot obtain accurate equality constraints for the parameters. Zhu and Huang [1] suggested that the equality constraints should be released in the filtering algorithm after a certain number of iterations, in order to avoid the problem introduced by inaccuracy of constraints. However, when to release the constraints remains difficult to determine.

We use inequality constraints in this work. Inequality constraints on parameters can be written as

$$d^L \leq c(\theta) \leq d^U \tag{1.10}$$

where d^L and d^U indicate lower and upper bounds of the inequality constraints. Inequality constraints are the most common and widely available relationships in practical applications. Once the inequality parameter constraints are specified, it is natural to incorporate them into filtering algorithms in order to achieve better estimation performance.

The main objective of this thesis is to develop efficient methods for constrained estimation with inequality parameter constraints. The methods are developed under the assumption that we have a deterministic model with Gaussian noise. The preliminary objective is to use different existing algorithms to estimate states and parameters simultaneously. Then, we need to develop an approach to obtain inequality parameter constraints from routine operation data. The ultimate goal of this thesis is to develop constrained estimation methods to obtain improved estimation performance. The efficacies of the proposed approaches will be demonstrated using various simulated process examples.

1.2 A brief literature review

The literature on estimation theory and its applications have received considerable attention over the last half century. It began from two major techniques, the Luenberger observer [2] and the celebrated Kalman filter [3]. For a deterministic model with no random noise, the Luenberger observer and its extensions can be used for time-invariant systems with known parameters. The equation for the Luenberger observer contains a term that corrects the current state estimate using a proportional gain to the prediction error. This correction term ensures the stability and convergence of the observer even when the system being observed is unstable. The Kalman filter provides the optimal estimation for a linear system corrupted with Gaussian noise. In the Kalman filter, there is also a correction gain called the Kalman gain to ensure stability and convergence. The theory behind the Kalman filter is well established and its applications have grown significantly in both academia and industry [4].

The most widely used approach in nonlinear state estimation is the extended Kalman filter (EKF). The EKF employs a first-order Taylor approximation to linearize the nonlinear model around the current state estimate. The EKF provides a suboptimal estimate of a nonlinear model with Gaussian noise. However, for a highly nonlinear system, a significant amount of linearization error will arise due to the covariance propagation and update which are carried out through the linearized model [5]. Therefore, it is necessary to investigate new nonlinear state estimation algorithms.

The unscented Kalman filter (UKF), which is developed by Julier et al. [6], is a better alternative compared to the EKF in handling nonlinear systems. The UKF is based on a deterministic sampling technique called the unscented transformation (UT), where a set of points representing the state distribution are chosen and propagated through the nonlinear model. The mean and covariance of the estimate are recovered from these points. The employment of the UT in the UKF results in more accurate capture of the mean and covariance in nonlinear propagation. The posterior mean and covariance estimated from the sample points are accurate to the second order for any nonlinearity [7]. If the priori random variable is Gaussian, the posterior mean and covariance are accurate to the third order for any nonlinearity [8].

The ensemble Kalman filter (EnKF), originally proposed by Evenson [9], is another approach for nonlinear estimation. Instead of the deterministic sampling strategy used in the UKF, the EnKF employs the Monte Carlo sampling method to generate a large number of random samples for carrying nonlinear prediction and update. Since the EnKF approximates the covariance terms by averaging over a large number of samples, it is expected to give better results as the ensemble size increases. This has been demonstrated in Gillijns et al. [10], where simulation results show a steady decrease in estimation errors as the ensemble size grows. The Kalman filter, the extended Kalman filter, the unscented Kalman filter, and the ensemble Kalman filter mentioned above are among the well established techniques that can be used for dynamic state estimation. All these filters have a Kalman filter structure, which follows a prediction-update procedure.

Besides the Kalman filter and its variants, there is another approach for state estimation called moving horizon estimation (MHE). MHE is based on the Bayesian *maximum a posteriori* approach and is an optimization based method for state estimation. The concept of MHE was originally proposed to overcome the limitations of the Kalman filter in handling of constraint and nonlinearity. In the MHE, the estimates are obtained by minimizing an objective function in which there are penalties for measurement error, state error and the arrival cost. The arrival cost is employed in the MHE to summarize the information before the moving window in order to reduce the computational load. Therefore, the MHE can explicitly use a set of measurements measured over a horizon with a certain length. This measurement horizon or measurement window is moving forward in time. Specifically, Haseltine and Rawlings [11] critically compared the performance of MHE to an extended Kalman filter (EKF) and concluded that MHE consistently provides improved state estimation and greater robustness to both poor guesses of the initial state and tuning parameters.

As mentioned earlier, we often have certain constraints (obtained from physical knowledge or analysis of operating data) that the estimated states and parameters need to satisfy. The only approach that naturally incorporates such constraints is the MHE. However, MHE has major drawbacks related to on-line implementation since it relies on solving a nonlinear programming problem at each sampling time. Also, the approximation of the arrival cost in constraint handling in MHE remains a difficult problem. A common approach to implement constraints in Kalman filters is known as clipping [11], where the estimate is projected onto the boundaries of the constrained space if it lies outside. Some strict equality constraint can lead to reduction in degrees of freedom (model reduction) in estimation problems [12]. Vachhani et al. [13] developed the recursive nonlinear dynamic data reconciliation (RNDDR) method, which combines the computational advantages of recursive estimation with constraint handling. Recently, López-Negrete et al. [14] have shown that constrained recursive particle filters can be used for estimating the arrival cost in MHE.

1.3 Thesis outline

The rest of the thesis is organized as follows.

In Chapter 2, the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) are presented and compared. The comparison shows that the UKF is superior to the EKF in parameter estimation. An approach to constructing inequality parameter constraints from steady-state routine operating data is introduced. A constrained parameter estimation scheme with the UKF and a constraint implementation method with inequality parameter constraints are introduced. The efficacy of the proposed method is demonstrated on a CSTR process as well as a PMMA polymerization reactor.

In Chapter 3, the ensemble Kalman filter (EnKF), which is based on the Monte Carlo sampling method, is presented. Its performance is assessed with different ensemble sizes, and is also compared to the UKF. The appropriate use of the projection method in constraining the particles in the EnKF is introduced. A new constrained parameter estimation method with the EnKF, which results in better performance than the projection method, is proposed. A CSTR process is employed in this chapter to demonstrate the performance of the proposed method.

In Chapter 4, we consider the moving horizon estimation (MHE) for constrained estimation. The unconstrained MHE algorithm is presented along with common approaches of arrival cost approximation. The problem with the MHE when incorporating constraints directly is addressed in this chapter. An alternative method for constrained parameter estimation with the MHE that provides better performance than directly constrained MHE is proposed. The efficacy of the proposed method is demonstrated on a CSTR process.

Chapter 5 gives conclusions and recommendations for future work.

Chapter 2

Inequality Constrained Parameter Estimation with Unscented Kalman Filter

Parameter estimation of a dynamic system is an important task in process systems engineering. The utilization of an augmented system offers the approach of estimating process states and parameters simultaneously. In practice, the true parameters often satisfy certain constraints which need to be incorporated into the estimation procedure. In this chapter, we consider the inequality constrained parameter estimation problem with the unscented Kalman filter (UKF). We first show that the UKF is superior to the extended Kalman filter (EKF) for parameter estimation problem. We then introduce a method of constructing inequality constraints for parameters from routine operating data; this offers a way to use steady-state data and its associated model to obtain constraints on parameters. We also propose a method of constraint implementation with the UKF that yields faster convergence than the conventional methods of projection or clipping. The application of the proposed constrained estimation method provides fast recovery of state and parameter estimates from inaccurate initial guesses; this is demonstrated on two continuous chemical processes with discrete measurements.

2.1 Introduction

Process control requires an accurate model characterizing the process in order to achieve good process monitoring, online optimization and control performance. A mathematical model usually includes a number of algebraic and differential equations which represent the process dynamics. In fact, most process models and measurements are corrupted by noise and errors, which results in inaccuracy of the sampled data. Estimation of unknown parameters from a set of available measurements is often a major goal in setting up a model. Moreover, the estimation of states is also essential for process control. Hence, the dual estimation problem naturally arises. Generally, an augmented system, where unknown parameters are augmented as states, is employed to deal with this problem. It offers an approach to estimate unknown parameters and states simultaneously. With the parameters as augmented states, the dual estimation problem becomes a nonlinear filtering problem. Various estimation algorithms exist to sequentially estimate a nonlinear system with online measurements.

Sequential filtering algorithms such as the extended Kalman filter (EKF), the unscented Kalman filter (UKF) [15] and moving horizon estimation (MHE) [16] are powerful tools for nonlinear state estimation. The most widely used algorithm for nonlinear systems is the EKF, which employs the Jacobians (first order approximations) to locally linearize the model so that the conventional Kalman filter (KF) algorithm can be applied. In the EKF, the nonlinear model can be used to compute the predicted state as well as the predicted output. However, the covariance cannot be updated directly through a nonlinear model; instead, a linearized model has to be used. Difficulties arise from its use of linearization [5]. The estimation performance with the EKF may not be desirable for highly nonlinear models due to its linearization error. Also, the calculation of Jacobian matrices can be difficult for high-dimensional systems. To overcome the difficulties encountered with EKF, Julier et al. [6] proposed the UKF. It uses the unscented transformation (UT), which employs a set of weighted points (called sigma points) to represent the estimate mean and covariance, to propagate the system nonlinearity. It has computational efficiency owing to its Kalman filtering structure, as well as a better approximation than the EKF for nonlinear systems. Furthermore, it eliminates the need for Jacobian calculation. The superior performance of the UKF when compared with the EFK is demonstrated by Wan and van der Merwe [17] and Romanenko and Castro [18]. MHE, which is an optimization based algorithm, can provide good estimation of nonlinear systems by solving a nonlinear programming (NLP) problem over a finite horizon. However, the computational efficiency remains an issue for a long horizon or a large number of decision variables [14].

For practical processes of interest, model parameters typically satisfy some constraints, either linear or nonlinear, equality or inequality constraints. For example, the surface area of a reactor should be positive (and fall within a reasonable range). In most cases, we can obtain inequality constraints that give bounds of the parameter to be estimated [1]. In this chapter, we introduce an approach to constructing inequality parameter constraints from noisy steady-state measurement data. Since we have the inequality constraints on parameters, a proper implementation method should be applied to incorporate the constraints into the filtering algorithm. Inequality constraints are naturally handled by moving horizon estimation (MHE) due to its optimization based algorithm. However, MHE requires a heavy on-line computational load and the exact arrival cost is hard to determine for the constrained estimation. Vachhani et al. [19] proposed the recursive nonlinear dynamic data reconciliation (RNDDR) method, in which the constraints are taken into consideration and the nonlinear state and covariance propagation are based on the EKF algorithm. Vachhani et al. [20] later proposed the unscented recursive nonlinear dynamic data reconciliation (URNDDR) method through a combination of the UKF and RNDDR, which gives more accurate and efficient estimation performance for nonlinear constrained estimation. Recently, Prakash et al. [21] proposed constrained ensemble Kalman filter (C-EnKF) [21] and constrained particle filter (C-PF) [22] with the use of constrained Monte Carlo samples and probability density function (PDF) truncation for nonlinear estimation. In this paper, we propose a new constraint handling method with the UKF for inequality constrained parameter estimation.

The rest of this chapter is organized as follows. In Section 2.2, we review and compare the EKF and the UKF filtering algorithms and show that the UKF is superior to the EKF in nonlinear estimation. In Section 2.3, an approach to constructing inequality parameter constraints from routine operating data is introduced. In Section 2.4, we propose a constrained parameter estimation scheme with inequality constraints, as well as a constraint implementation method which provides better performance than the conventional projection method. Finally, the proposed method is demonstrated on a CSTR process and a PMMA polymerization reactor system to show the improvement both on the estimation performance and on the control performance.

2.2 Extended Kalman filter and unscented Kalman filter

The Kalman filter is limited to linear systems; however, it is the basis for many estimators for nonlinear systems, including the extended Kalman filter (EKF), the unscented Kalman filter (UKF) and the ensemble Kalman filter (EnKF). In this section, we investigate the performance of the EKF and the UKF for parameter estimation.

2.2.1 Extended Kalman filter

The extended Kalman filter is the most common approach for nonlinear state estimation. It is a linearized version of the Kalman filter and provides suboptimal estimation. The following continuous-discrete nonlinear model is considered,

$$\dot{x} = f(x, u) + w \tag{2.1a}$$

$$y_k = h(x_k) + v_k \tag{2.1b}$$

In the EKF, the nonlinearities of the systems are approximated by first-order Taylor expansions. Consider the linearization of the nonlinear state transition function f in Eq. 2.1a around the nominal state x_n:

$$f(x, u) \approx f(x_n, u) + \left.\frac{\partial f}{\partial x}\right|_{x_n} (x - x_n) \tag{2.2}$$

Also, the measurement transition function h in Eq. 2.1b can be linearized as

$$h(x) \approx h(x_n) + \left.\frac{\partial h}{\partial x}\right|_{x_n} (x - x_n) \tag{2.3}$$

In the EKF algorithm, the most recent estimate is used as the nominal state. For the stochastic nonlinear model described in Eq. 2.1, the EKF algorithm, which can be divided into two groups, *prediction* and *update*, is summarized below [4].

- **Prediction**

$$\dot{\hat{x}}(t)^- = f(\hat{x}(t), u(t)) \tag{2.4}$$

$$\dot{P}(t)^- = F(\hat{x}(t))P(t) + P(t)F^T(\hat{x}(t)) + Q \tag{2.5}$$

- **Update**

$$K_k = P_k^- H(\hat{x}_k^-)^T (H(\hat{x}_k^-) P_k^- H^T(\hat{x}_k^-) + R)^{-1} \tag{2.6}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - h(\hat{x}_k^-)) \tag{2.7}$$

$$P_k = (I - K_k H(\hat{x}_k^-)) P_k^- \tag{2.8}$$

$$F(\hat{x}(t)) \equiv \left.\frac{\partial f}{\partial x}\right|_{\hat{x}(t)}, \qquad H(\hat{x}_k^-) \equiv \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_k^-}$$

where $\hat{x}^-$ denotes the prior state estimate, and P and P^- denote the posterior estimate covariance and prior estimate covariance respectively. In the prediction step, the propagation of the state estimate and covariance to the next time step via the process model in Eq. 2.1a gives the prior state estimate and covariance. In the update step, a posterior state estimate and covariance is obtained using a correction based on the current available observation y_k. K is called the Kalman gain that updates the prior estimate to the posterior estimate using the error innovation term $(y_k - H\hat{x}_k^-)$ at each time instant k. The EKF makes use of the noisy available data y on a linearized system with Gaussian noise to continuously update the current state estimate of the system.

It is important to state that the extended Kalman filter is in general not an optimal filter. The EKF typically works well only when the first-order Taylor linearization adequately approximates the nonlinear function [4]. For a highly nonlinear state function f and transition function h, the estimation performance may not be adequate due to the linearization error. Another concern for application of the EKF is in the initialization; when the initial guess may be far from the true value, the filter may diverge quickly [11].

2.2.2 Unscented Kalman filter

In nonlinear state estimation with the EKF, the covariance is propagated through the linearization of the nonlinear model. Therefore, the estimated covariance matrix tends to represent the true covariance poorly, and estimation becomes inconsistent in the statistical sense. The idea of the unscented Kalman filter (UKF) proposed by Julier et al. [15, 6] comes from trying to capture the state estimate and covariance more accurately in the nonlinear sequential estimation problem.

The main technique involved in the UKF is known as the unscented transformation (UT), which is a method for the nonlinear transformation of the mean and the covariance in filters and estimators [17]. In the UT, a deterministic set of carefully chosen points, called sigma points $\mathcal{X}$, are used to capture the true mean and covariance of a state estimate $\hat{x}$. When sigma points are propagated through the true nonlinear functions (f and h), the posterior mean and covariance are adequately captured. The unscented transformation is illustrated in Fig.2.1.

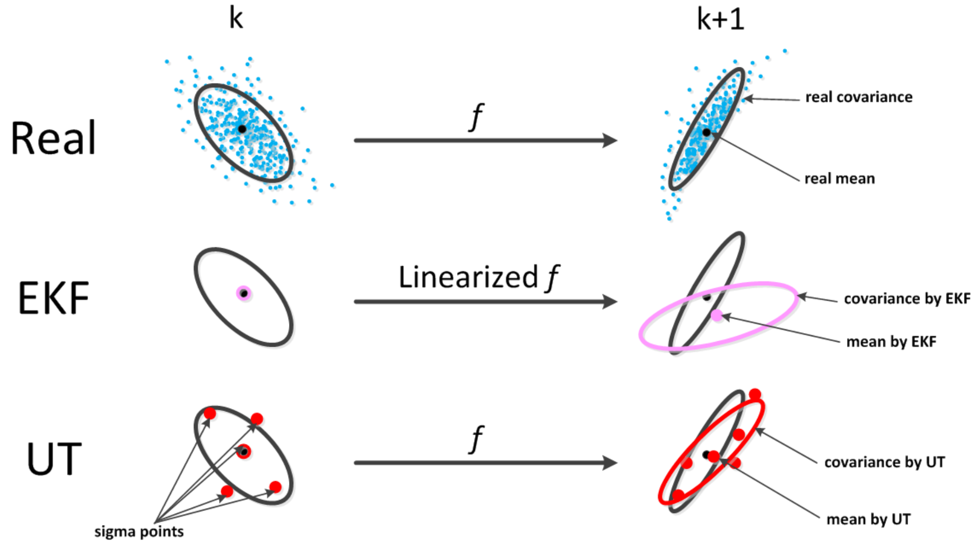


Figure 2.1: Schematic of the unscented transformation (UT) for mean and covariance propagation, and comparison with the linearized approach of the EKF

The top figure in Fig.2.1 shows the actual mean and covariance propagation, the middle figure shows the EKF approach, and the bottom figure shows the unscented transformation (UT) approach. It is clear that the UT is superior to the EKF in nonlinear approximation [17].

The unscented Kalman filter is a recursive nonlinear estimation algorithm based on the UT. A summary of the UKF algorithm is given below [4].

- **Selection of Sigma Points**

$$\hat{x}_k^a = \begin{bmatrix} \hat{x}_k \\ 0 \\ 0 \end{bmatrix}, \qquad P_k^a = \begin{bmatrix} P_k & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{bmatrix}$$

$$\mathcal{X} = \left[\hat{x}_k^a \quad \hat{x}_k^a + \gamma\sqrt{P_k^a} \quad \hat{x}_k^a - \gamma\sqrt{P_k^a}\right] \tag{2.9}$$

- **Prediction**

$$\mathcal{X}^- = f(\mathcal{X}, u) \tag{2.10}$$

$$\mathcal{Y} = h(\mathcal{X}) \tag{2.11}$$

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_i^m \mathcal{X}^- \tag{2.12}$$

$$\hat{y}_k = \sum_{i=0}^{2L} W_i^m \mathcal{Y} \tag{2.13}$$

$$P_k^- = \sum_{i=0}^{2L} W_i^c \left[\mathcal{X}^- - \hat{x}_k^-\right]\left[\mathcal{X}^- - \hat{x}_k^-\right]^T \tag{2.14}$$

- **Update**

$$P^{yy} = \sum_{i=0}^{2L} W_i^c \left[\mathcal{Y} - \hat{y}_k\right] \left[\mathcal{Y} - \hat{y}_k\right]^T \tag{2.15}$$

$$P^{xy} = \sum_{i=0}^{2L} W_i^c \left[\mathcal{X}^- - \hat{x}_k^-\right] \left[\mathcal{Y} - \hat{y}_k\right]^T \tag{2.16}$$

$$K_k = P^{xy} P^{yy-1} \tag{2.17}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - \hat{y}_k) \tag{2.18}$$

$$P_k = P_k^- - K_k P^{yy} K_k^T \tag{2.19}$$

where L is the dimension of state x_k^a, $\lambda = \alpha^2(L + \kappa) - L$ and $\gamma = \sqrt{(L + \lambda)}$. W^m and W^c are weights for calculating the mean and covariance of the $2L + 1$ vectors in $\mathcal{X}^-$ and $\mathcal{Y}$, which are generated by Eq. 2.10 and Eq. 2.11 respectively. Weights are given by

$$W_0^m = \frac{\lambda}{L + \lambda} \tag{2.20}$$

$$W_0^c = \frac{\lambda}{L + \lambda} + (1 - \alpha^2 + \beta) \tag{2.21}$$

$$W_i^m = W_i^c = \frac{1}{2(L + \lambda)} \qquad i = 1, \cdots, 2L \tag{2.22}$$

where α, β, κ are parameters involved in the UKF (see [23] for details).

In the unscented Kalman filter, the nonlinearities are approximated by nonlinear transformation of sigma points, which returns much better estimation performance than the first-order Taylor expansion (Jacobian) of the nonlinear model in the EKF. In addition, the UKF removes the requirement of calculating the Jacobian matrices F and H and is more computationally efficient.

Comparison between the EKF and the UKF for parameter estimation

It is stated above that the UKF is superior to the EKF and is easier to use for nonlinear estimation problems. In this section, a fast electrical circuit system is employed to assess the parameter estimation performance with the EKF and the UKF.

The continuous-discrete circuit system is shown in Eq. 2.23 [1]. u is the system input, and θ_1 and θ_2 are system parameters with true values of 1.2 and 0.0015, respectively. Both state and measurement equations involve Gaussian noise, and available measurements y are recorded discretely with sampling time $T_s = 0.0002s$.

$$\dot{x} = \frac{-x}{\theta_1 \times \theta_2} + \frac{u}{\theta_2} + w \tag{2.23a}$$

$$y_k = x_k + v_k \tag{2.23b}$$

θ_1 and θ_2 are of interest and assumed unknown in the above system. In order to estimate the state x and the parameters θ_1 and θ_2 simultaneously, we have the augmented nonlinear model formulated as

$$\dot{x}_1 = \frac{-x_1}{0.0015 \times x_2} + \frac{u}{0.0015} + w_1 \tag{2.24a}$$

$$\dot{x}_2 = 0 + w_2 \tag{2.24b}$$

$$\dot{x}_3 = 0 + w_3 \tag{2.24c}$$

$$y_k = x_{1,k} + v_k \tag{2.24d}$$

Parameter θ_1 and θ_2 are augmented as states x_2 and x_3 respectively, and are modeled as random walk processes shown in Eq. 2.24b and Eq. 2.24c. The initial guesses of parameters θ_1 and θ_2 are 1 and 0.1 respectively. Covariance matrices for the process and measurement zero-mean Gaussian noise are set as

$$Q = \begin{bmatrix} 1 \times 10^{-5} & 0 & 0 \\ 0 & 1 \times 10^{-5} & 0 \\ 0 & 0 & 1 \times 10^{-5} \end{bmatrix}, \qquad R = 1 \times 10^{-4} \tag{2.25}$$

A comparison of the estimation performance for parameter θ_1 between the EKF and the UKF is shown in Fig.2.2. The result shows the superior performance of the UKF, which has faster convergence for estimation of parameter than EKF; this is in agreement with the theoretical background of their algorithms.

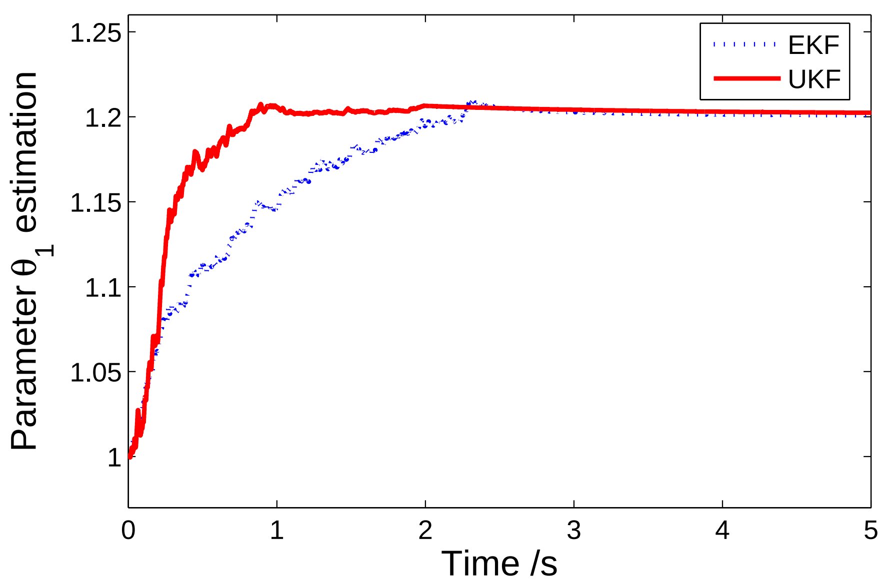


Figure 2.2: Comparison of the performance between the EKF and the UKF for the estimation of parameter θ_1 of the continuous-discrete electrical circuit system

2.3 Construction of inequality parameter constraints

The inequality parameter constraints $d^L \leq c(\theta) \leq d^U$ will be considered in this thesis. In this section, a method of constructing inequality constraints on parameters from routine operation data is introduced. The constraints given in Eq.1.10 can be obtained from some steady-state measurements, say $\{y_1, y_2, \ldots, y_N\}$, from the process of interest. In order to do that, we need to relate the measurements and parameters together using the form given below

$$y = c(\theta) + e, e \sim \mathcal{N}(0, \sigma_e) \tag{2.26}$$

where e is assumed as Gaussian noise with standard deviation σ_e. Therefore, given a equation like Eq. 2.26, a confidence interval (inequality constraints) of $c(\theta)$ can be calculated by conventional statistical methods. Hence, the inequality constraint on the parameter, with approximately 99.99% confidence, can be obtained as

$$[\hat{c}(\theta) - 4\sigma(\hat{c}(\theta)), \hat{c}(\theta) + 4\sigma(\hat{c}(\theta))]$$

where $\hat{c}(\theta)$ and $\sigma(\hat{c}(\theta)$ are estimates of the parameter relationship $c(\theta)$ and its standard deviation, respectively.

The following two sections show the procedure of obtaining inequality parameter constraints from available steady-state measurements $\{y_1, y_2, \ldots, y_N\}$ for linear systems and nonlinear systems respectively.

2.3.1 Linear system

The linear continuous-discrete system is given as

$$\dot{x}(t) = Ax(t) + Bu(t) + w \tag{2.27}$$

$$y_k = Hx_k + v_k \tag{2.28}$$

For this linear system at steady-state operating condition, meaning $\dot{x}$ equals 0. It gives

$$\dot{x} = Ax + Bu + w = 0 \tag{2.29}$$

Rewrite Eq.2.29 gives

$$x = -A^{-1}Bu - A^{-1}w \tag{2.30}$$

Substituting Eq. 2.30 into Eq. 2.28 gives (the subscript k has been omitted for ease of notation)

$$y = -HA^{-1}Bu - HA^{-1}w + v \tag{2.31}$$

In Eq. 2.31, y is the set of steady-state measurements of the linear system and u is a known input of the system. Therefore, we can treat $-HA^{-1}Bu$ as the parameter relationship $c(\theta)$. Since w and v are Gaussian, $-HA^{-1}w + v$ is written as Gaussian noise e. Thus, Eq. 2.31 can be written in the form of Eq. 2.26

$$\begin{aligned} y &= \underbrace{-HA^{-1}Bu} \qquad \underbrace{-HA^{-1}w + v} \\ y &= \quad c(\theta) \qquad + \qquad e \end{aligned} \tag{2.32}$$

As long as we have N steady-state measurements $\{y_1, y_2, \ldots, y_N\}$, we can obtain the inequality constraints on parameters, $d^L \leq c(\theta) \leq d^U$, by conventional statistical method. The calculation procedure is shown as follows:

- Calculate $\hat{c}(\theta)$

$$\hat{c}(\theta) = \frac{y_1 + y_2 + \ldots + y_N}{N} \tag{2.33}$$

- Calculate $Var(\hat{c}(\theta))$

$$Var(\hat{c}(\theta)) = Var(\frac{y_1 + y_2 + \ldots + y_N}{N}) = \frac{{\sigma_e}^2}{N} \tag{2.34}$$

${\sigma_e}^2$ can be approximated by the sample variance of y, ${S_y}^2$. i.e.,

$$ {\sigma_e}^2 \approx {S_y}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{2.35}$$

where $\bar{y}$ denotes the sample mean: $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$.

Thus, we have

$$Var(\hat{c}(\theta)) \approx \frac{{S_y}^2}{N} \tag{2.36}$$

- Calculate $\sigma(\hat{c}(\theta))$

$$\sigma(\hat{c}(\theta)) = \sqrt{Var(\hat{c}(\theta))} \approx \frac{S_y}{\sqrt{N}} \tag{2.37}$$

- Finally, we have the 99.99% confidence interval for $c(\theta)$ as

$$[\hat{c}(\theta) - 4\sigma(\hat{c}(\theta)), \hat{c}(\theta) + 4\sigma(\hat{c}(\theta))]$$

The above procedure is easy to use for inequality constraints calculation from sampled steady-state data.

2.3.2 Nonlinear system

Most systems encountered in the real world are nonlinear, and nonlinear models are required to achieve adequate modeling accuracy in many industrial processes. The problem considered in this section is the calculation of the inequality constraints of $c(\theta)$ for a nonlinear stochastic system based on available noisy steady-state measurements. Consider the following nonlinear stochastic model

$$\dot{x} = f(x, u, \theta) + w \tag{2.38a}$$

$$y_k = h(x_k) + v_k \tag{2.38b}$$

where w and v_k are zero-mean Gaussian noises.

At steady-state, we have $\dot{x} = 0$, that is

$$\dot{x} = f(x, u, \theta) + w = 0 \tag{2.39}$$

The goal of our work, again, is the construction of the relationship between measurements y and parameters θ, without having state x involved (because x is unknown). We linearize nonlinear equations Eq. 2.38b and Eq. 2.39 to eliminate the variable x. The linearization of Eq. 2.38b and Eq. 2.39 gives

$$y = h(x_{ss}) + H(x_{ss})(x - x_{ss}) + v \tag{2.40}$$

$$f(x_{ss}, u, \theta) + F(x_{ss})(x - x_{ss}) + w = 0 \tag{2.41}$$

$$F(x_{ss}) \equiv \left.\frac{\partial f}{\partial x}\right|_{x_{ss}}, \qquad H(x_{ss}) \equiv \left.\frac{\partial h}{\partial x}\right|_{x_{ss}}$$

where x_{ss} denotes the steady-state point around which the nonlinear functions are linearized. Because x is a common term in Eq. 2.40 and Eq. 2.41, it can be eliminated by combining these two equations. Eq. 2.41 can be further written as

$$x = -F(x_{ss})^{-1} f(x_{ss}, u, \theta) - F(x_{ss})^{-1} w + x_{ss} \tag{2.42}$$

Substituting Eq. 2.42 into Eq. 2.40 gives

$$y = h(x_{ss}) - H(x_{ss})F(x_{ss})^{-1} f(x_{ss}, u, \theta) - H(x_{ss})F(x_{ss})^{-1} w + v \tag{2.43}$$

In Eq. 2.43, y is the steady-state measurement, and $h(x_{ss}) - H(x_{ss})F(x_{ss})^{-1} f(x_{ss}, \theta)$ is the parameter relationship $c(\theta)$. $-H(x_{ss})F(x_{ss})^{-1} w + v$ is Gaussian noise denoted by e. Thus, Eq. 2.43 can be written in form of Eq. 2.26 as well.

$$\begin{aligned} y &= \underbrace{h(x_{ss}) - H(x_{ss})F(x_{ss})^{-1} f(x_{ss}, u, \theta)} \qquad \underbrace{-H(x_{ss})F(x_{ss})^{-1} w + v} \\ y &= \qquad\qquad c(\theta) \qquad\qquad + \qquad\qquad e \end{aligned} \tag{2.44}$$

Then, the inequality constraint for $c(\theta)$ can be calculated using the same statistical method as introduced above in Section 2.3.1.

2.3.3 Illustrative example

In this section, we apply the developed method to a linear system to demonstrate the construction of the inequality constraint on parameters from available steady-state data. Let us again consider the linear electrical circuit system shown in Eq. 2.23.

$$\dot{x} = \frac{-x}{\theta_1 \times \theta_2} + \frac{u}{\theta_2} + w \tag{2.23a}$$

$$y_k = x_k + v_k \tag{2.23b}$$

For an input, say $u = 0.9$, applied to the electrical circuit system, we can collect a number of steady-state measurements, say 50, $\{y_1, y_2, \ldots, y_{50}\}$. The noisy measurements are shown in Fig.2.3.

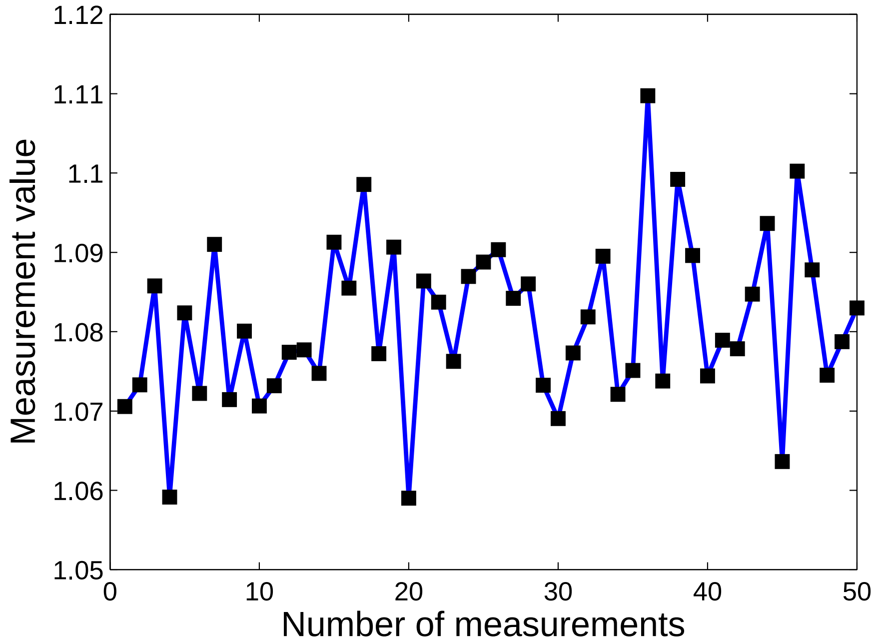


Figure 2.3: Plot of steady state measurements of the continuous-discrete electrical circuit system

Once we have these steady-state measurements, the method proposed in Section 2.3.1 is applied to develop the inequality parameter constraint. For this linear electrical circuit system, the parameter relationship $c(\theta)$ can be obtained through Eq. 2.32 as follows

$$c(\theta) = -HA^{-1}Bu = \theta_1 u = 0.9\theta_1 \tag{2.46}$$

Therefore, we have the equation between measurements y and parameter relationship $c(\theta)$ shown as

$$y = 0.9\theta_1 + e, \qquad e \sim \mathcal{N}(0, \sigma_e) \tag{2.47}$$

The inequality constraint for $c(\theta)$, which is $0.9\theta_1$, is calculated as follows:

- Calculate $\hat{c}(\theta)$

$$\hat{c}(\theta) = \frac{y_1 + y_2 + \ldots + y_{50}}{50} = 1.08105 \tag{2.48}$$

- Calculate $\sigma(\hat{c}(\theta))$

$$\sigma(\hat{c}(\theta)) = \sqrt{Var(\hat{c}(\theta))} \approx \frac{1}{\sqrt{50}} \cdot \sqrt{\frac{1}{50-1}\sum_{i=1}^{50}(y_i - \bar{y})^2} = 0.00145 \tag{2.49}$$

  where $\bar{y}$ denotes sample mean: $\bar{y} = \frac{1}{50}\sum_{i=1}^{50} y_i = 1.08105$.

- Then, the 99.99% confidence interval of $c(\theta)$ is obtained as

$$[\hat{c}(\theta) - 4\sigma(\hat{c}(\theta)), \hat{c}(\theta) + 4\sigma(\hat{c}(\theta))] = [1.0752, 1.0869] \tag{2.50}$$

Thus, we have 99.99% confidence that the inequality parameter constraint is

$$1.0752 \leq 0.9\theta_1 \leq 1.0869 \tag{2.51}$$

2.4 Inequality constrained parameter estimation with the UKF

Constrained estimation incorporates more information than merely using available measurements in the recursive filtering algorithm. In Section 2.2, we have shown that the UKF is superior to the EKF in parameter estimation. In Section 2.3, we show that it is appropriate and useful to use inequality constraints in estimation. In this section, we first introduce the overall scheme of constrained estimation using the UKF and inequality parameter constraints. This recursive constrained estimation scheme provides improved estimation compared to the estimation without considering constraints. Secondly, we propose a method of inequality constraints implementation.

2.4.1 Constrained estimation framework

The proposed constrained estimation scheme is shown in Fig.2.4.

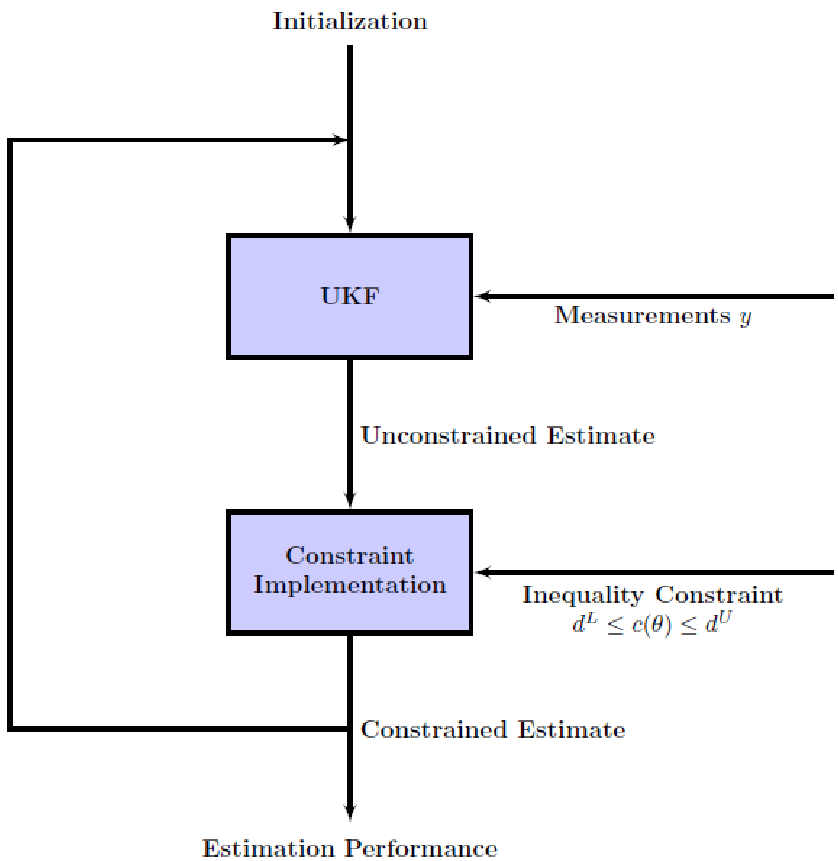


Figure 2.4: Framework of the proposed constrained estimation scheme

As shown in Fig.2.4, we use both the knowledge of available measurements and inequality constraints to achieve constrained parameter estimation. With available measurements, the unconstrained estimate will be obtained using the UKF. After that, the unconstrained estimate will be further processed with the knowledge of inequality constraints to obtain a constrained estimate. The constrained estimation scheme retains the recursive nature of the estimator.

2.4.2 Implementation of constraints

Recall that in the parameter estimation problem, the state vector x and parameter vector θ are combined into an augmented state x_a (Eq. 1.4):

$$x_a = \begin{bmatrix} x \\ \theta \end{bmatrix} \tag{5}$$

With the inequality constraints shown in Eq. 1.10, the constraints are on the parameters θ rather than on the state x. In the following content, $\tilde{x}$ and $\tilde{\theta}$ denote the constrained state and parameter estimates respectively.

In order to transform the unconstrained estimate to a constrained one, different implementation approaches have been developed and presented in the literature [19, 20, 21, 22, 24]. In the following discussion, we will review two common implementation methods, then propose a new method.

Existing implementation methods

1. Projection method

The projection method is a basic and intuitive approach of implementing constraints. It projects the violating estimates on to the constraint boundaries. The constrained estimate $\tilde{x}_a$ can be obtained by solving the following optimization problem

$$\tilde{x}_a = argmin_{x_a}(x_a - \hat{x}_a)^T W(x_a - \hat{x}_a) \tag{2.52}$$

subject to

$$d^L \leq c(\theta) \leq d^U$$

where $\hat{x}_a$ is the unconstrained estimate. Various approaches can be employed to solve this constrained quadratic programming problem. In this work, we use `fmincon` in MATLAB to obtain the constrained estimate.

In the projection method, since deviations from the unconstrained estimates $\hat{x}_a$ are penalized, the constrained estimate will tend to lie on the constraint boundary. Moreover, the estimate will only be changed in the parameter θ because of the inequality parameter constraints. This method can be illustrated as

$$\hat{x}_a = \begin{bmatrix} \hat{x} \\ \hat{\theta} \end{bmatrix} \longrightarrow \tilde{x}_a = \begin{bmatrix} \hat{x} \\ \tilde{\theta} \end{bmatrix}$$

2. Recursive nonlinear dynamic data reconciliation method (RNDDR)

The recursive nonlinear dynamic data reconciliation (RNDDR) method developed by Vachhani et al. [13] also enables the incorporation of constraints into recursive estimation. This method not only constrains the estimates, but also provides the additional advantage of minimizing the measurement error. The constrained estimate $\tilde{x}_a$ is calculated by solving

$$\tilde{x}_a = argmin_{x_a}(y - h(x_a))^T R^{-1}(y - h(x_a)) + (x_a - \hat{x}_a)^T P^{-1}(x_a - \hat{x}_a) \qquad (2.53)$$

such that

$$d^L \leq c(\theta) \leq d^U$$

In solving the above objective function, the estimate could be changed both in the state x and the parameter θ:

$$\hat{x}_a = \begin{bmatrix} \hat{x} \\ \hat{\theta} \end{bmatrix} \longrightarrow \tilde{x}_a = \begin{bmatrix} \tilde{x} \\ \tilde{\theta} \end{bmatrix}$$

When used with the unconstrained UKF, this method is called the URNDDR [20]. This method fails mainly because the first term in the objective function is not even needed, because of the fact that y is only a function of x as shown in Eq.2.1b. Thus it will result in the inconsistency within the constrained state estimate $\tilde{x}$ and constrained parameter estimate $\tilde{\theta}$. An example where this strategy fails will be presented later in Section 2.5.1.

Proposed method

As mentioned earlier, the projection method can be used to obtain the constrained parameter estimate without changing the state estimate. The nonlinear data reconciliation method, which has a plausible perspective of trying to minimize the measurement error, may not work properly mainly because that the first term in the objective function of Eq.2.53 does not depends on parameter θ. In order to constrain the parameter estimate as well as to minimize the measurement error, we propose a new method to implement the inequality parameter constraints, in which the constrained estimate is obtained by solving the following objective function

$$\tilde{x}_{ak} = argmin_{x_{ak}}(y_k - \hat{y}_k)^T R^{-1}(y_k - \hat{y}_k) + (x_{ak} - \hat{x}_{ak})^T P^{-1}(x_{ak} - \hat{x}_{ak}) \qquad (2.54)$$

subject to

$$d^L \leq c(\theta) \leq d^U$$

where $\hat{y}_k$ is the estimated output propagated from $\begin{bmatrix} \hat{x}_{k-1} \\ \theta \end{bmatrix}$ through the nonlinear function f and h. Therefore, Eq. 2.54 can be further written as

$$\tilde{x}_{ak} = argmin_{x_{ak}}(y_k - h(f(\begin{bmatrix} \hat{x}_{k-1} \\ \theta \end{bmatrix})))^T R^{-1}(y_k - h(f(\begin{bmatrix} \hat{x}_{k-1} \\ \theta \end{bmatrix}))) + (x_{ak} - \hat{x}_{ak})^T P^{-1}(x_{ak} - \hat{x}_{ak}) \qquad (2.55)$$

The main difference between RNDDR method and proposed method is that parameters θ is the only decision variable in the latter. By solving the optimization problem, an optimal parameter θ that falls into the constraints can be obtained. The progression of the estimate is illustrated by

$$\hat{x}_a = \begin{bmatrix} \hat{x} \\ \hat{\theta} \end{bmatrix} \longrightarrow \tilde{x}_a = \begin{bmatrix} \hat{x} \\ \tilde{\theta} \end{bmatrix}$$

The essence of this method is to utilize the parameter constraints to update the parameter estimate, in other words, the process model will be more precisely recovered; hence, we can expect a better estimate in turn in the filtering step.

As mentioned above, the state estimate at time $k-1$ is employed to predict $\hat{y}_k$ in this method. In all recursive filtering algorithms such as the Kalman filter, the EKF or the UKF, the state estimate at time $k-1$ is obtained using available measurements up to time $k-1$, i.e., $\hat{x}_{k-1|k-1}$. As long as we have measurements at time k on hand, it is more appropriate to use a smoothed state estimate, which uses information up to time k, i.e., $\hat{x}_{k-1|k}$, to predict $\hat{y}_k$. In the following content, we use $\widehat{xs}_{k-1}$ to denote the smoothed estimate $\hat{x}_{k-1|k}$.

The Kalman smoother, which is a backward recursive filtering algorithm, provides the optimal state estimate at time $k-1$ using information up to time $k+N$. The smoothed estimate $\widehat{xs}_{k-1}$ can be calculated using the following Kalman smoother equation

$$\widehat{xs}_{k-1} = \hat{x}_{k-1} + L(\hat{x}_k - \hat{x}_k^-) \tag{2.56}$$

where L is the Kalman smoother gain. For a linear model, L is calculated as follows

$$L = P_{k-1}A^T(P_k^-)^{-1} \tag{2.57}$$

For a nonlinear model, A could be the Jacobian matrix of the nonlinear state equation f for approximation. For more accurate smoothing of nonlinear models, a smoother based on the unscented transformation (UT) called the unscented Rauch-Tung-Striebel smoother (URTSS) may be employed [25]. In URTSS, L is calculated using

$$L = D(P_k^-)^{-1} \tag{2.58}$$

where

$$D = \sum W_i^c \left[\mathcal{X}_{k-1} - \hat{x}_{k-1}\right]\left[\mathcal{X}_k^- - \hat{x}_k^-\right]^T$$

$\mathcal{X}$ and W_i^c have the same meaning as in the UKF (see Section 2.2).

2.5 Simulation examples

2.5.1 Example 1: CSTR process

In this section, we apply the proposed algorithm to a dynamic chemical process for demonstrating the performance of the proposed inequality constrained parameter estimation algorithm. The example is a continuous stirred tank reactor (CSTR) where a first-order, irreversible (A→B), exothermic ($\Delta H < 0$) reaction between A and B takes place [26]. The model equations are as follows

$$\dot{C}_A = \frac{q}{V}(C_{Af} - C_A) - k_0 e^{-\frac{E}{RT}} C_A \tag{2.59a}$$

$$\dot{T} = \frac{q}{V}(T_f - T) + \frac{(-\Delta H)}{\rho C_p} k_0 e^{-\frac{E}{RT}} C_A + \frac{UAr}{V\rho C_p}(T_c - T) \tag{2.59b}$$

where C_A, q and T are the concentration, flow rate and temperature inside the reactor respectively. $(\cdot)_f$ denotes the feeding. V is the reactor volume, k_0 is the pre-exponential factor, E is the activation energy, R is the universal gas constant, ρ

is the liquid density in CSTR, C_p is the heat capacity, ΔH is the enthalpy change of the reaction, U and Ar are the heat transfer coefficient and area between the CSTR and the jacket respectively, and T_c is the temperature of the cooling jacket.

The nonlinear model given by Eq. 3.17 has two states, concentration C_A (state x_1) and temperature T (state x_2) and one manipulated variable T_c (input u). In this case study, we assume ρ and $\frac{E}{R}$ are the parameters to be estimated, while their true values are 1000 and 8750 respectively. The values of all the parameters in the CSTR process are specified in Table 2.1 [26].

Table 2.1: Parameters in the CSTR model

Variable	Value	Variable	Notation
C_{Af}	1 mol/L	C_A	x_1
q	100 L/min	T	x_2
T_f	350 K	T_c	u
V	100 L	ρ	θ_1
k_0	7.2×10^{10} min^{-1}	$\frac{E}{R}$	θ_2
C_p	0.239 J/g·K		
ΔH	-5×10^4 J/mol		
UAr	5×10^4 J/min·K		

With the parameter values specified in Table 2.1, we have the following model:

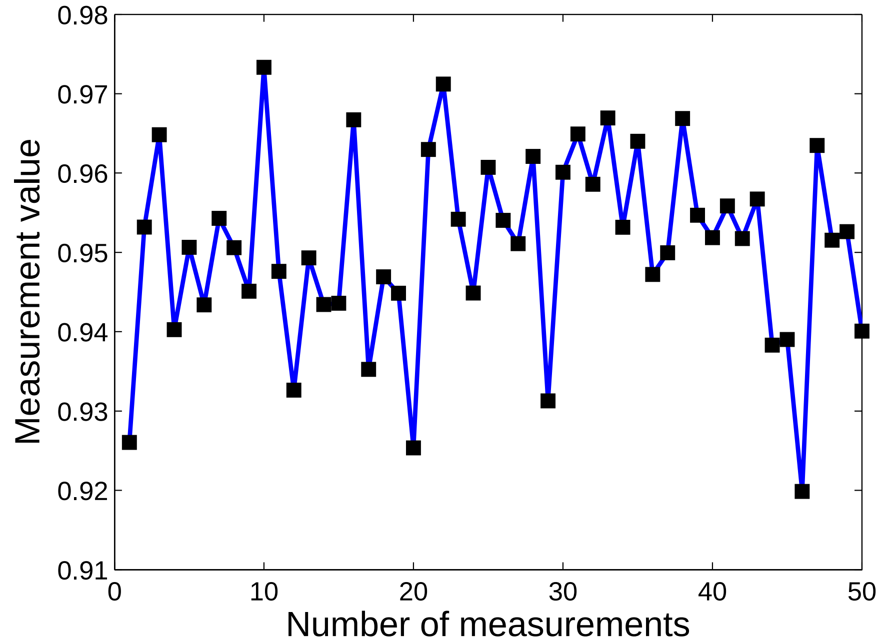
$$\dot{x}_1 = 1 - x_1 - 7.2 \times 10^{10} e^{-\frac{\theta_2}{x_2}} x_1 + w_1 \tag{2.60}$$

$$\dot{x}_2 = 350 - x_2 + 150.6276 \times 10^{14} e^{-\frac{\theta_2}{x_2}} \frac{x_1}{\theta_1} + 2092.05 \frac{(u - x_2)}{\theta_1} + w_2 \tag{2.61}$$

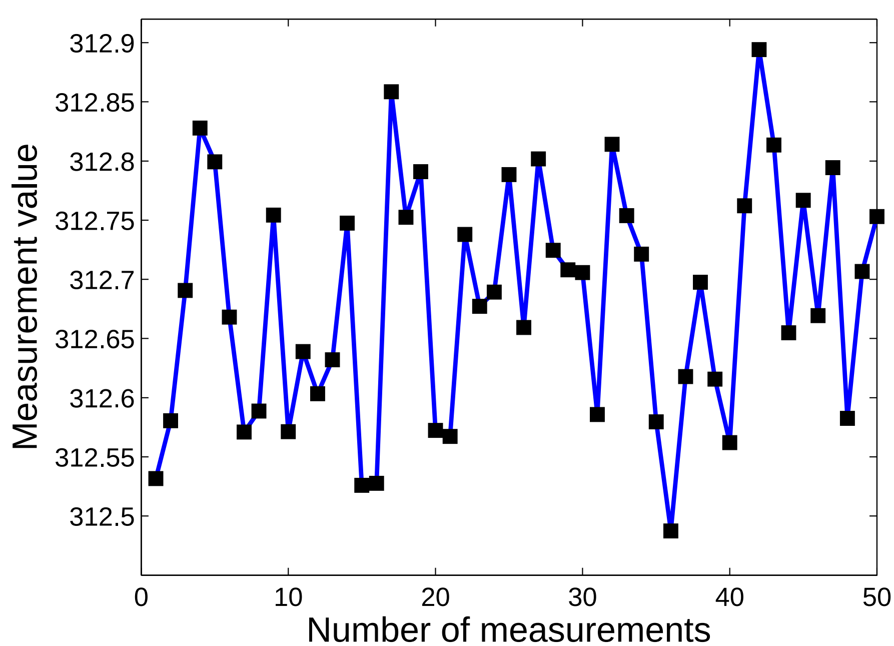
$$y_{1,k} = x_{1,k} + v_{1,k} \tag{2.62}$$

$$y_{2,k} = x_{2,k} + v_{2,k} \tag{2.63}$$

Applying an input, say $u = 290$ K, to the system, we can collect a number of measurements $\{y_{1,1}, y_{1,2}, \ldots, y_{1,N}\}$ and $\{y_{2,1}, y_{2,2}, \ldots, y_{2,N}\}$ after the system reaches steady-state. For example, we have 50 steady-state measurements of y_1 and y_2 respectively, which are plotted in Fig.2.5.



(a) Plot of steady-state measurements y_1 of the CSTR model.



(b) Plot of steady-state measurements y_2 of the CSTR model.

Figure 2.5: Plot of steady-state measurements of the CSTR model.

For the nonlinear CSTR model, the method proposed in Section 2.3.2 can be applied to construct the inequality parameter constraints with these available steady-state measurements.

Eq. 2.60 and Eq. 2.61 can be expressed in form of a general nonlinear equation as

$$\dot{x} = f(x, u, \theta) + w \tag{2.64}$$

Eq. 2.62 and Eq. 2.63 can be rewritten in the same way as

$$y = h(x) + v \tag{2.65}$$

It is shown in Eq. 2.44 that the parameter relationship for nonlinear system is constructed as

$$c(\theta) = h(x_{ss}) - H(x_{ss})F(x_{ss})^{-1}f(x_{ss}, u, \theta) \tag{2.66}$$

where $F(x_{ss})$ and $H(x_{ss})$ are Jacobian matrices defined as

$$F(x_{ss}) \equiv \left.\frac{\partial f}{\partial x}\right|_{x_{ss}}, \qquad H(x_{ss}) \equiv \left.\frac{\partial h}{\partial x}\right|_{x_{ss}}$$

x_{ss} is the steady-state point around which the nonlinear functions are linearized, and we can use the mean value of measurements y as the steady-state point in this case.

$$x_{ss} = \text{mean of } y = \begin{bmatrix} 0.9510 \\ 312.66 \end{bmatrix} \tag{2.67}$$

Once x_{ss} is obtained, we can derive the following expression as

$$\begin{aligned}
f(x_{ss}, u, \theta) &= \begin{bmatrix} 0.049 - 6.8472 \times 10^{10} e^{-\frac{\theta_2}{312.6621}} \\ 37.3379 + \frac{1.4325\times 10^{16}}{\theta_1} e^{-\frac{\theta_2}{312.6621}} - \frac{4.741\times 10^{4}}{\theta_1} \end{bmatrix} \\
F(x_{ss}) &= \begin{bmatrix} -1 - 7.2 \times 10^{10} e^{-\frac{\theta_2}{312.6621}} & -7.0043 \times 10^{5} e^{-\frac{\theta_2}{312.6621}} \theta_2 \\ \frac{1.5063\times 10^{16}}{\theta_1} e^{-\frac{\theta_2}{312.6621}} & -1 + 1.4653 \times 10^{11} e^{-\frac{\theta_2}{312.6621}} \frac{\theta_2}{\theta_1} - \frac{2092.05}{\theta_1} \end{bmatrix} \\
h(x_{ss}) &= \begin{bmatrix} 0.9510 \\ 312.66 \end{bmatrix} \\
H(x_{ss}) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\end{aligned}$$

The parameter relationship $c(\theta)$ can be calculated using Eq. 2.66 as follows

$$c(\theta) = \begin{bmatrix} c_1(\theta) \\ c_2(\theta) \end{bmatrix} = h(x_{ss}) - H(x_{ss})F(x_{ss})^{-1}f(x_{ss}, u, \theta) \tag{2.68}$$

Therefore, we have the equations built between measurement y and parameter relationship $c(\theta)$ as

$$y_1 = c_1(\theta) + e_1, \qquad e_1 \sim \mathcal{N}(0, \sigma_{e_1}) \tag{2.69}$$
$$y_2 = c_2(\theta) + e_2, \qquad e_2 \sim \mathcal{N}(0, \sigma_{e_2}) \tag{2.70}$$

From the 50 steady-state measurements y_1 and y_2 shown in Fig.2.5, the 99.99% confidence intervals of $c(\theta)$ can be obtained using the method introduced in Section 2.3.1, and are shown below as

$$[\hat{c}_1(\theta) - 4\sigma(\hat{c}_1(\theta)), \hat{c}_1(\theta) + 4\sigma(\hat{c}_1(\theta))] = [0.9451, 0.9570]$$
$$[\hat{c}_2(\theta) - 4\sigma(\hat{c}_2(\theta)), \hat{c}_2(\theta) + 4\sigma(\hat{c}_2(\theta))] = [312.6003, 312.7239]$$

Thus, we have 99.99% confidence that the inequality parameter constraints are

$$\begin{bmatrix} 0.9451 \\ 312.6003 \end{bmatrix} \leq \begin{bmatrix} c_1(\theta) \\ c_2(\theta) \end{bmatrix} \leq \begin{bmatrix} 0.9570 \\ 312.7239 \end{bmatrix} \tag{2.71}$$

where the expressions for $c_1(\theta)$ and $c_1(\theta)$ are obtained through Eq. 2.68.

Since we have the above inequality parameter constraints, constrained estimation can be performed based on the scheme shown in Fig.2.4. In the nonlinear filtering part, the UKF is employed to obtain the optimal unconstrained estimate. The unknown parameters θ_1 and θ_2 are treated as states x_3 and x_4, which are modeled as random walk processes shown in Eq. 2.72c and Eq. 2.72d. With the available discrete noisy measurements of concentration x_1 and temperature x_2, we have the overall continuous-discrete nonlinear system formulated as

$$\dot{x}_1 = 1 - x_1 - 7.2 \times 10^{10} e^{-\frac{x_4}{x_2}} x_1 + w_1 \tag{2.72a}$$

$$\dot{x}_2 = 350 - x_2 + 150.6276 \times 10^{14} e^{-\frac{x_4}{x_2}} \frac{x_1}{x_3} + 2092.05 \frac{(u - x_2)}{x_3} + w_2 \tag{2.72b}$$

$$\dot{x}_3 = 0 + w_3 \tag{2.72c}$$

$$\dot{x}_4 = 0 + w_4 \tag{2.72d}$$

$$y_{1,k} = x_{1,k} + v_{1,k} \tag{2.72e}$$

$$y_{2,k} = x_{2,k} + v_{2,k} \tag{2.72f}$$

The initial guesses for θ_1 and θ_2 are 1025 and 8755 respectively. The process and measurement Gaussian noise covariance for the system are set as

$$Q = \begin{bmatrix} 1 \times 10^{-5} & 0 & 0 & 0 \\ 0 & 1 \times 10^{-3} & 0 & 0 \\ 0 & 0 & 1 \times 10^{-5} & 0 \\ 0 & 0 & 0 & 1 \times 10^{-5} \end{bmatrix}, \qquad R = \begin{bmatrix} 1 \times 10^{-4} & 0 \\ 0 & 1 \times 10^{-2} \end{bmatrix}$$

After we obtain the unconstrained estimate from the UKF, the inequality parameter constraints in Eq. 2.71 are employed to obtain the constrained estimate with different implementation algorithms. The constrained estimation scheme, as shown in Fig.2.4, runs recursively to estimate the states and parameters.

Fig.2.6 shows the comparison between the projection method and the RNDDR method of parameter ρ in the CSTR model. In Fig.2.6, both estimators have the same initial condition ($\rho = 1025$), while the true value is 1000. The simulation result shows that the projection method gives the correct estimation result while the RNDDR method leads to a failure in estimating the true value of the parameter.

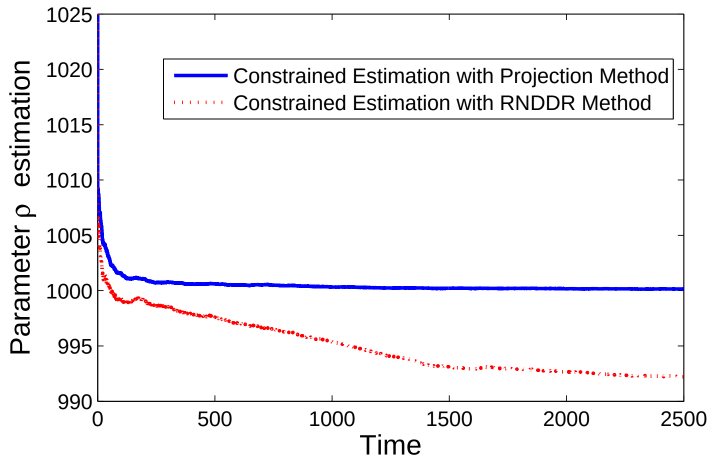


Figure 2.6: Simulation results showing the failure of RNDDR in the estimation of parameter ρ of the CSTR model.

The proposed method is then applied. As discussed in Section 2.4.2, the smoothed estimate $\widehat{xs}_{k-1}$ should provide better performance than $\hat{x}_{k-1}$ in the application of proposed method. The comparison is again illustrated on the estimation of parameter ρ in the CSTR model. Fig.2.7 shows the estimation error of parameter ρ by using different state estimate at time $k-1$, with both estimators having the same initial condition. The simulation result shows a significant improvement in parameter estimation by employing the smoothed estimate $\widehat{xs}_{k-1}$.

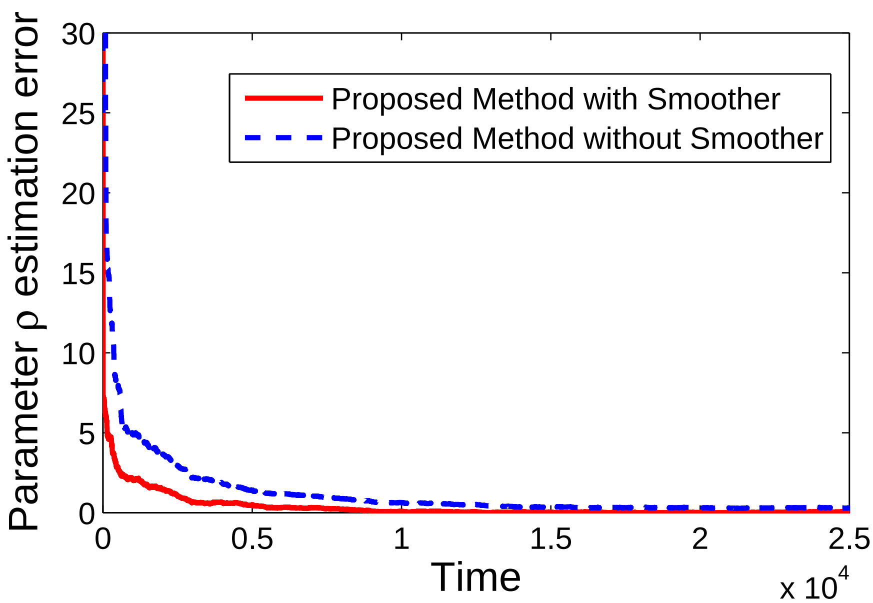


Figure 2.7: Simulation results showing the advantage of employing smoothed estimate $\hat{x}_{k-1|k}$ for estimation of parameter ρ of the CSTR model.

Fig.2.8 and Fig.2.9 show the overall simulation comparison for the estimation of CSTR model between unconstrained estimation, constrained estimation by the projection method and constrained estimation by the proposed method. It is seen that inequality constrained estimation can significantly improve the parameter estimation performance. Furthermore, the proposed method can achieve better performance than the conventional projection method in constrained parameter estimation.

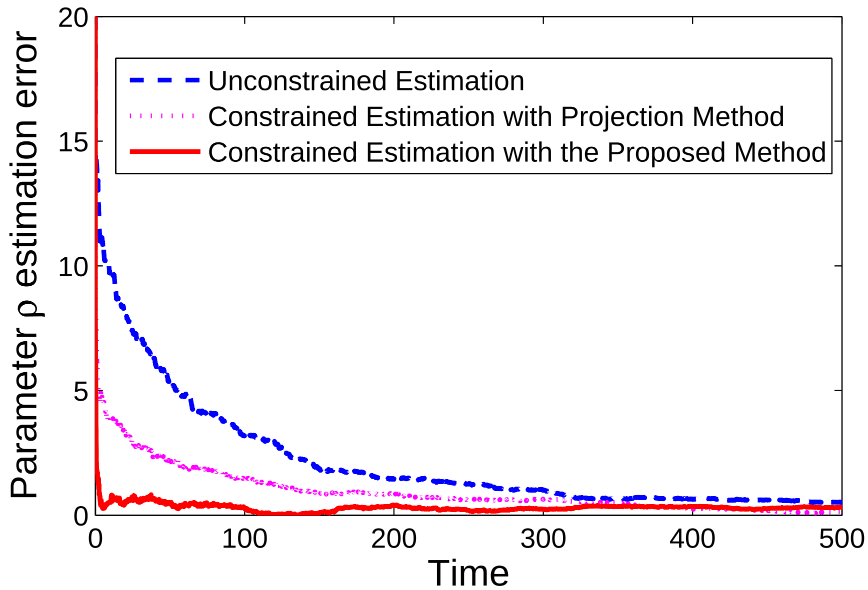


Figure 2.8: Simulation results of parameter ρ estimation error in the CSTR model.

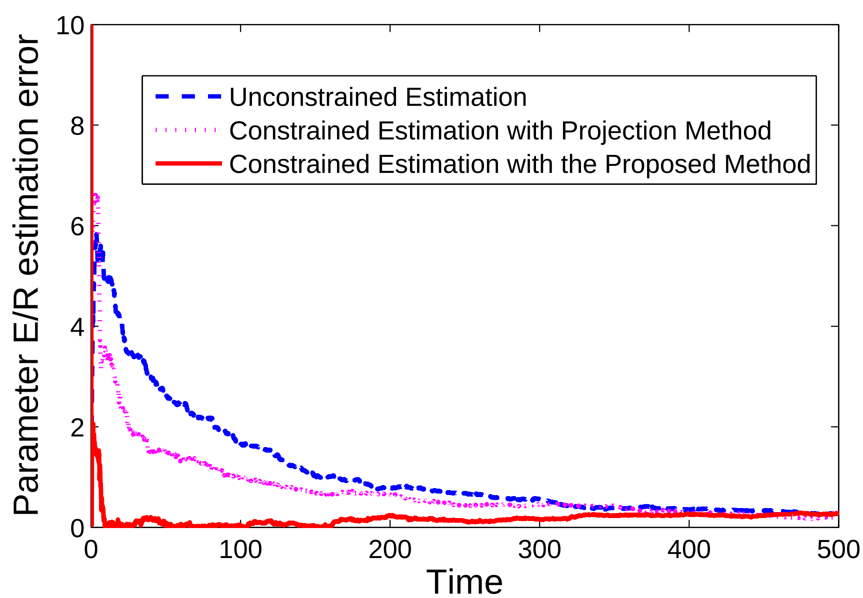


Figure 2.9: Simulation results for parameter $\frac{E}{R}$ estimation error in the CSTR model.

2.5.2 Example 2: PMMA polymerization reactor system

In this section, we employ a poly(methyl methacrylate) (PMMA) polymerization reactor (shown in Fig.2.10) for demonstrating the performance of the proposed inequality constrained parameter estimation algorithm. The process is described by the following differential equations [27]:

$$\frac{dC_m}{dt} = -\left[k_p \exp\left(\frac{-E_p}{RT}\right) + k_{f_m} \exp\left(\frac{-E_{f_m}}{RT}\right)\right] C_m P_0(C_I, T) + \frac{F(C_{m_{in}} - C_m)}{V} \tag{2.73a}$$

$$\frac{dC_I}{dt} = -k_I \exp\left(\frac{-E_I}{RT}\right) C_I + \frac{F_I C_{I_{in}} - F C_I}{V} \tag{2.73b}$$

$$\frac{dT}{dt} = k_p \exp\left(\frac{-E_p}{RT}\right) C_m \frac{(-\Delta H_p)}{\rho c_p} P_0(C_I, T) - \frac{UA}{\rho c_p V}(T - T_j) + \frac{F(T_{in} - T)}{V} \tag{2.73c}$$

$$\frac{dD_0}{dt} = \left[0.5 k_{T_c} \exp\left(\frac{-E_{T_c}}{RT}\right) + k_{T_d} \exp\left(\frac{-E_{T_d}}{RT}\right)\right] [P_0(C_I, T)]^2 + k_{f_m} \exp\left(\frac{-E_{f_m}}{RT}\right) C_m P_0(C_I, T) - \frac{F D_0}{V} \tag{2.73d}$$

$$\frac{dD_I}{dt} = M_m \left[k_p \exp\left(\frac{-E_p}{RT}\right) + k_{f_m} \exp\left(\frac{-E_{f_m}}{RT}\right)\right] C_m P_0(C_I, T) - \frac{F D_I}{V} \tag{2.73e}$$

$$\frac{dT_j}{dt} = \frac{F_{cw}}{V_o}(T_{w_o} - T_j) + \frac{UA}{\rho_w c_w V_o}(T - T_j) \tag{2.73f}$$

where

$$P_0(C_I, T) = \left[\frac{2 f^* C_I k_I \exp\left(\frac{-E_I}{RT}\right)}{k_{T_d} \exp\left(\frac{-E_{T_d}}{RT}\right) + k_{T_c} \exp\left(\frac{-E_{T_c}}{RT}\right)}\right]^{0.5} \tag{2.74}$$

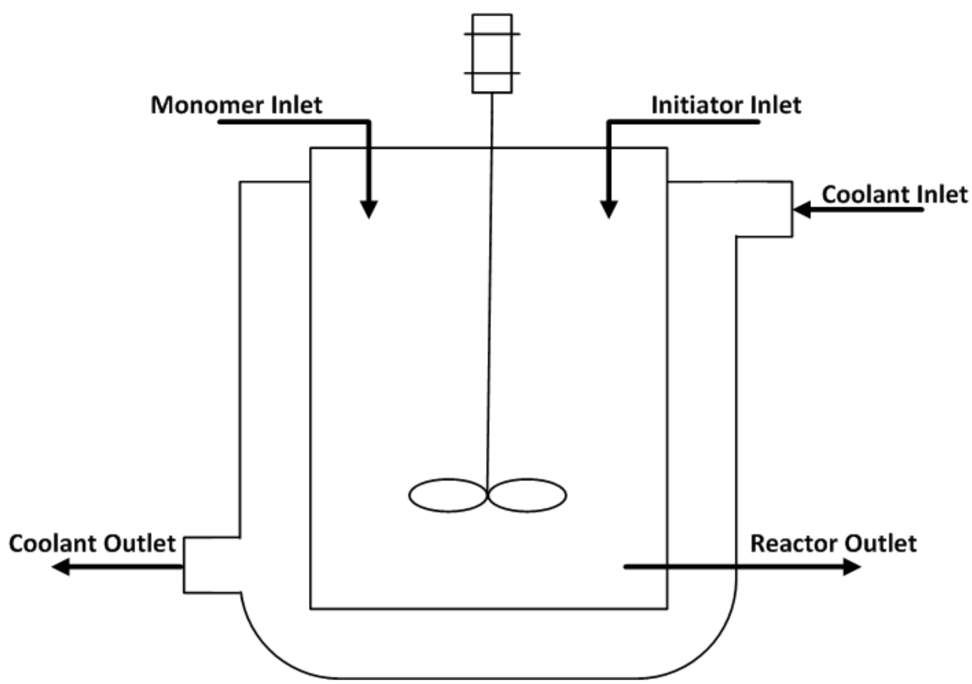


Figure 2.10: Schematic of poly(methyl methacrylate) (PMMA) polymerization reactor

Two measurements are available for this process:

$$y_1 = T + v_1 \tag{2.75a}$$
$$y_2 = T_j + v_2 \tag{2.75b}$$

where T_s is 0.013h.

In operating and controlling the PMMA polymerization reactor, the flowrate of the coolant F_{cw} is the manipulated variable u. The values of all the parameters involved in the reactor are listed in Table 2.2, and greater detail can be found in [27].

Table 2.2: Values of parameters in PMMA polymerization reactor system

Parameter	Value	Parameter	Value
k_p	1.77×10^9 kmol/m^3·h	E_p	1.83×10^4 kJ/kmol
k_{fm}	1.01×10^{15} kmol/m^3·h	E_{fm}	7.45×10^4 kJ/kmol
k_I	3.79×10^{18} kmol/m^3·h	E_I	1.29×10^5 kJ/kmol
k_{T_c}	3.82×10^{10} kmol/m^3·h	E_{T_c}	2.94×10^3 kJ/kmol
k_{T_d}	3.15×10^{11} kmol/m^3·h	E_{T_d}	2.94×10^3 kJ/kmol
R	8.314 kJ/kmol·K	F	1 m^3/h
$C_{m_{in}}$	8 kmol/m^3	V	0.1 m^3
F_I	0.017 m^3/h	$C_{I_{in}}$	6 kmol/m^3
$-\Delta H_p$	57800 kJ/kmol	ρ	866 kg/m^3
c_p	2 kJ/kg·K	U	720 kJ/h·K·m^2
A	2 m^2	T_{in}	350 K
M_m	100 kg/kmol	V_o	0.02 m^3
T_{w_o}	293 K	ρ_w	1000 kg/m^3
c_w	4.2 kJ/kg·K	f^*	0.58

In this parameter estimation problem, values of U and $-\Delta H$ are assumed unknown. As shown earlier, inequality parameter constraints can be obtained based on available steady-state measurements. Then, we use the available measurements y and inequality constraints to obtain the parameter and state estimates simultaneously. Note that the states D_0 and D_I are not observable from the measurements of T and T_j. Initial parameter guesses for the estimator are $U = 710$ and $-\Delta H = 57600$, while their true values are 720 and 57800 respectively. The covariance matrices of state noise and measurement noise are assumed to be

$$Q = diag\left[10^{-8}, 10^{-8}, 10^{-4}, 10^{-4}, 10^{-4}, 1\right], \qquad R = diag\left[10^{-3}, 10^{-3}\right]$$

Fig.2.11 and Fig.2.12 show the parameter estimation results on the PMMA polymerization reactor model. It is shown in Fig.2.11 that the proposed method has faster convergence for the parameter U than the projection method. As shown in Fig.2.12, the two methods have a similar performance for parameter $-\Delta H$. This is because the inequality constraints do not have much information constraining $-\Delta H$. In other words, parameter $-\Delta H$ has little impact on the steady-state measurements of the process.

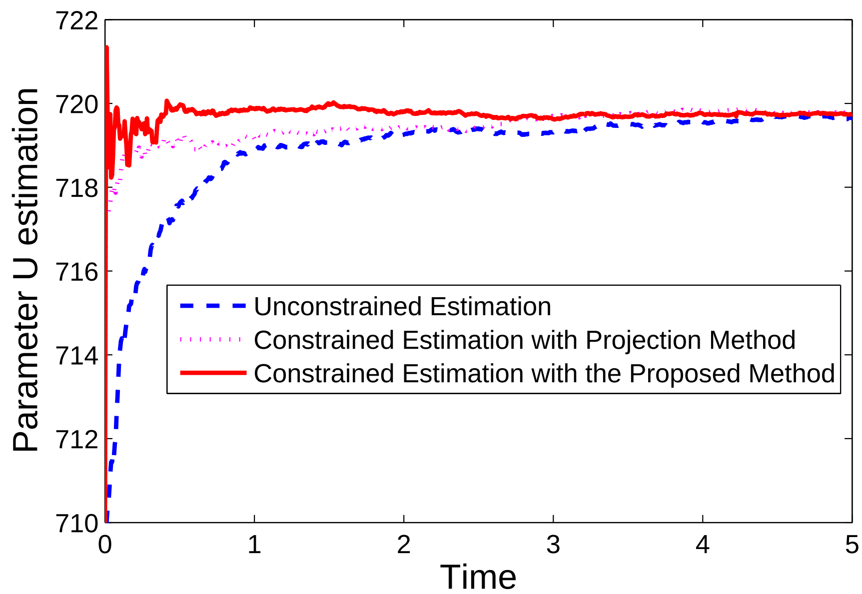


Figure 2.11: Simulation results for estimation of parameter U of the PMMA polymerization reactor model.

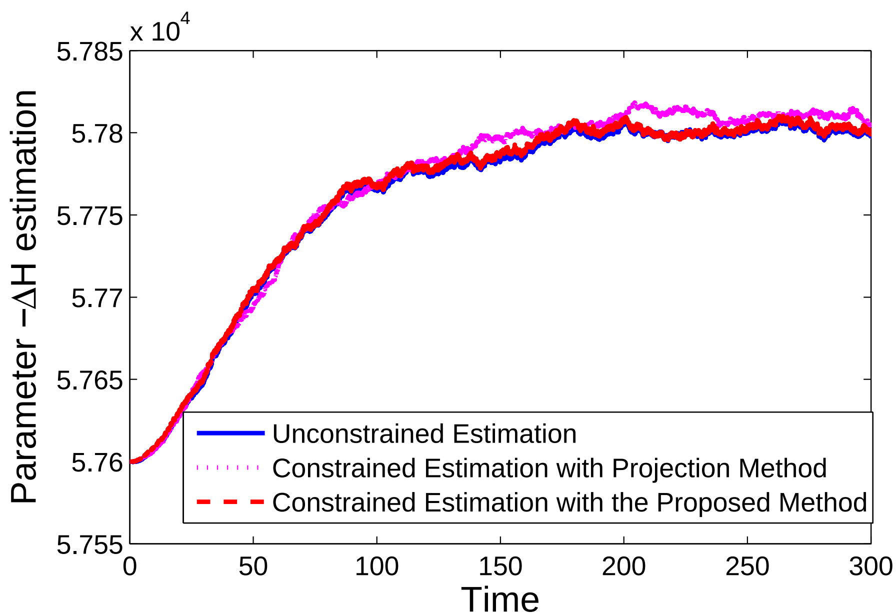


Figure 2.12: Simulation results for estimation of parameter $-\Delta H$ of the PMMA polymerization reactor model.

2.5.3 Examples on control improvement with estimators

Improvement of control performance is often the ultimate goal of state estimation. The estimators (filters), as introduced before, can be used to reduce the noise in the measurements, as well as to obtain the estimates of the unmeasurable variables. In this section, two control examples will be shown to illustrate the potential benefits of estimators in practical control applications.

The block diagram of a feedback control loop is shown in Fig.2.13.

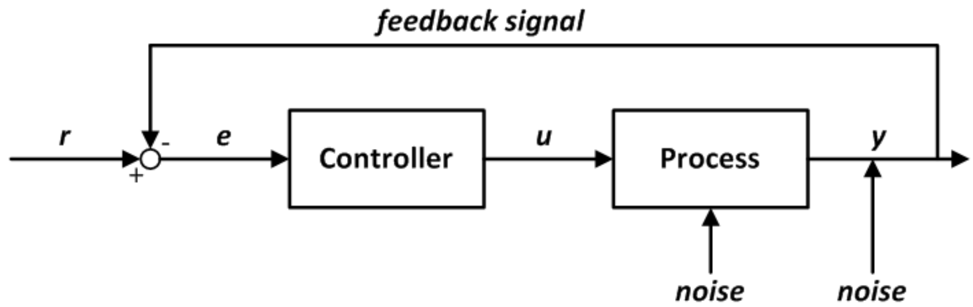


Figure 2.13: Block diagram of a feedback control loop

In Fig.2.13, the controller is employed for tracking the reference signal r. The main purpose of a tracking system is to keep a variable at a specified operating point by proper action of the controller, so that the plant can run smoothly and efficiently. The feedback signal normally comes from the measurements y. There is noise both in the process and in the measurement, which is assumed to be Gaussian in the simulations. For the tracking system shown in Fig.2.13, it can only be used to make the variables that are measurable to track the setpoint.

In Fig.2.14, an estimator is added into the tracking control system. For the measurable states, the estimator can be used to reduce the noise. For the unmeasurable states, they can be recovered by the estimator, so that they can also be controlled in an inferential manner.

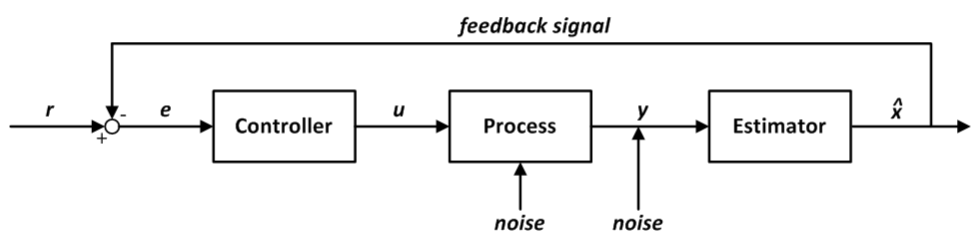


Figure 2.14: Block diagram of a feedback control loop with an estimator

In the following sections, estimation based tracking systems are employed for both the CSTR process and the PMMA polymerization reactor to illustrate the efficacies of estimators in control applications.

CSTR process

The CSTR process, as shown in Section 2.5.1, has two states C_A and T, which are both measurable with y_1 and y_2 respectively. The temperature of the cooling jacket T_c is the control input u. In this example, the operating temperature T (state x_2) is the variable to control. Assume that the CSTR process operates at a nominal condition $T = 324.48$ K. A PI controller with proportional gain of 3, integral gain of 0.08 is employed to change the operating condition from $T = 324.48$ K to $T = 282$ K. The control diagrams for this CSTR temperature tracking system are shown in Fig.2.15, where Fig.2.15(B) has an estimator.

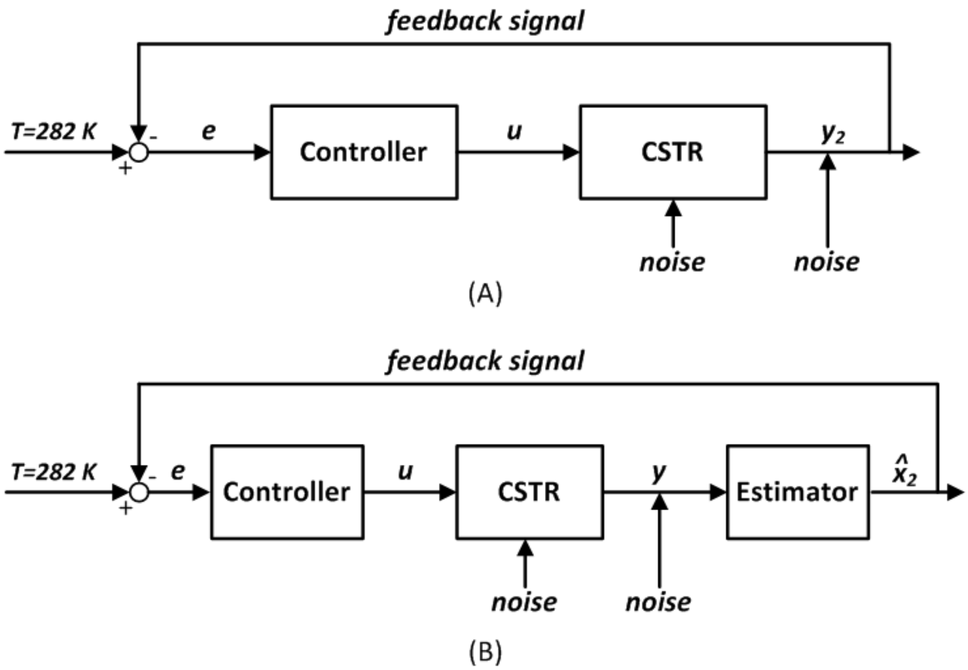


Figure 2.15: Block diagrams of temperature tracking systems for a CSTR process. (A) Feedback control; (B) Feedback control with estimator.

The comparison of tracking performance between control systems (A) and (B) is shown in Fig.2.16. The fluctuation around the reference signal comes from the noise in the measured output signal. It is shown that the estimator based control system has improved performance with much less fluctuation in the tracking of temperature T in this CSTR process.

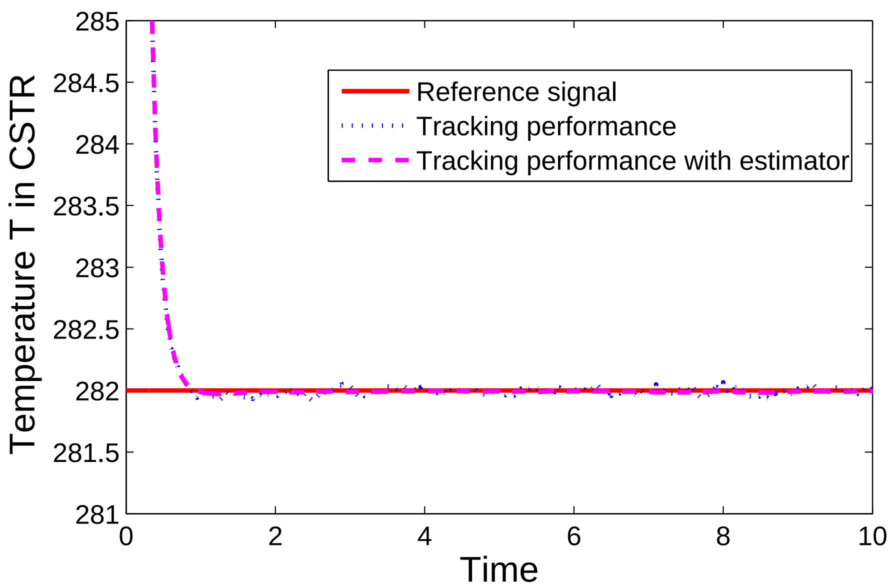


Figure 2.16: Performance comparison of tracking temperature T in a CSTR process

PMMA polymerization reactor

The six states of the PMMA polymerization reactor (shown in Section 2.5.2) are listed in Table 2.3.

Table 2.3: States of the PMMA polymerization reactor

State	Symbol	
x_1	C_m	observable, interest of tracking
x_2	C_I	observable
x_3	T	measurement y_1
x_4	D_0	unobservable
x_5	D_I	unobservable
x_6	T_j	measurement y_2

The concentration of monomer C_m in the reactor is of control interest as it indicates the quality of the polymer. Because the state C_m is not directly accessible through measurements, an estimator has to be employed in order to control the value of C_m. The control system diagram is shown in Fig.2.17, where a PI controller with proportional gain of 15, integral gain of 1 is used in the feedback control system.

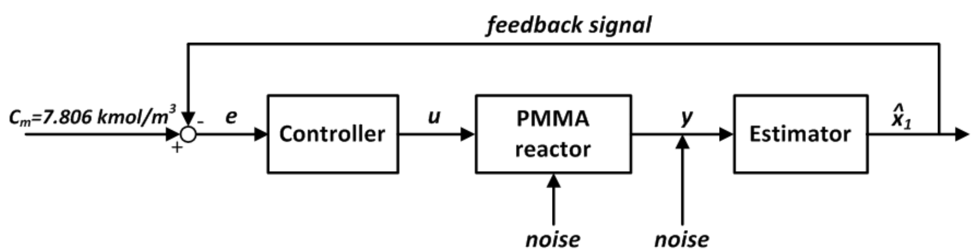


Figure 2.17: Block diagram of concentration tracking system for a PMMA polymerization reactor

Assume that the process is now operating with C_m = 7.73 kmol/m^3, while a reference signal with C_m = 7.8 kmol/m^3 is applied to the control system. The flowrate of the coolant F_{cw} is the control input u in this system.

In this control example, the values of the states and the parameters are unknown in the estimator. The initial guess of C_m is 7 kmol/m^3 while initial guesses for parameter U and$-\Delta H$ are 700 and 57400 respectively. Fig.2.18 shows the control performance with unconstrained estimator. It can be observed that there is a severe overshoot in the tracking performance in Fig.2.18. This overshoot is caused by the inaccurate estimate of x_1, which leads to the improper action of the controller.

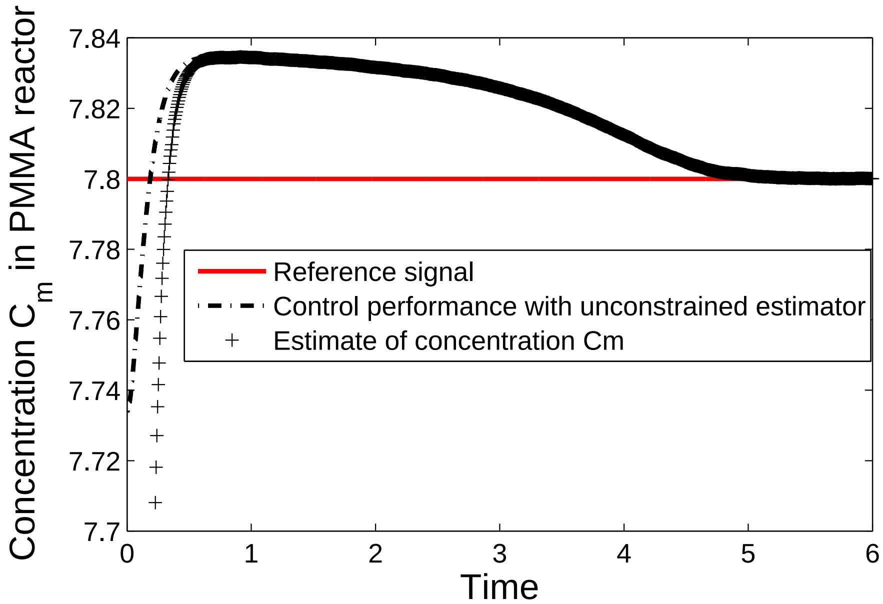


Figure 2.18: Control performance of tracking monomer concentration C_m with unconstrained estimator

Then, we use the constrained estimator in this control system. Parameter constraints can improve the state estimation performance, and consequently the control performance. Fig.2.19 shows the control performance with constrained estimator with inequality parameter constraints. The initial condition of the estimator is the same as for the unconstrained estimator. It is seen that the tracking performance with constrained estimator in Fig.2.19 is much better than the performance with unconstrained estimator in Fig.2.18, because of the faster recovery of the state estimation.

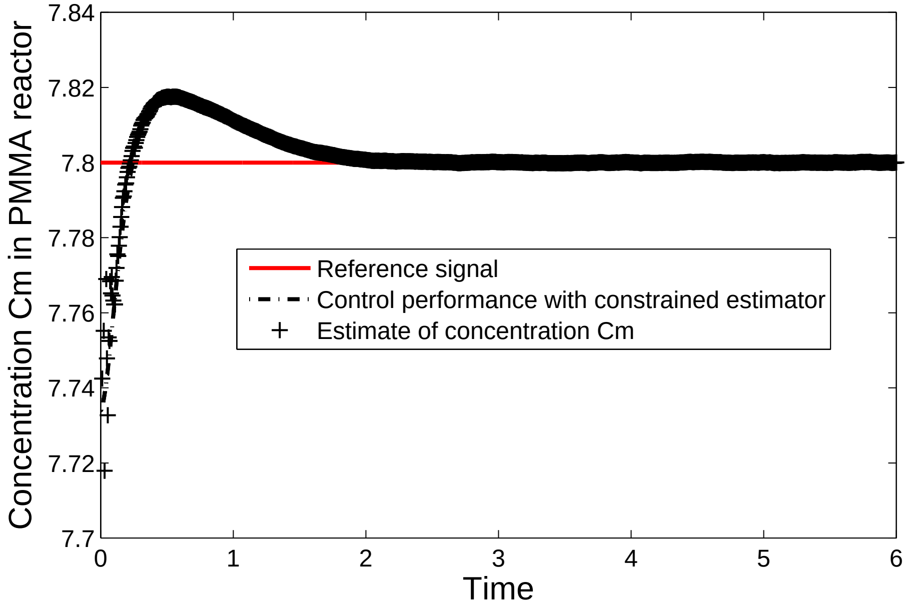


Figure 2.19: Control performance of tracking monomer concentration C_m with constrained estimator

2.6 Conclusions

The inequality constrained parameter estimation with the UKF has been considered in this chapter. In the comparison of parameter estimation with nonlinear filtering algorithms EKF and UKF, it is shown that the UKF has better performance than the EKF due to its more accurate description of the propagation of the covariance in nonlinear models. Next, we address the potential benefit of inequality parameter constraints to the estimation performance. An approach to constructing inequality parameter constraints for linear and nonlinear systems is developed, from which we can obtain constraints using steady-state routine operating data.

An inequality constrained estimation scheme is then proposed, as well as a constraint implementation method. The resulting algorithm constrains the parameter estimate and minimizes the measurement error. Moreover, it provides faster estimation convergence for parameters. Finally, the performance of the proposed inequality parameter constrained estimation method is demonstrated on a continuous-discrete CSTR process as well as a PMMA polymerization reactor system.

Chapter 3

Inequality Constrained Parameter Estimation with Ensemble Kalman Filter

Constrained parameter estimation with the ensemble Kalman filter (EnKF) is considered in this chapter. We first present the Monte Carlo sampling strategy for representing the distribution of the state estimate. Then, the EnKF algorithm is introduced. Its performance is assessed with different ensemble sizes and is also compared to the UKF. We then propose a projection method for constraining the particles in the EnKF without modifying the unconstrained covariance. This projection method can provide convergence in the constrained parameter estimation. We also propose a new constrained parameter estimation method with the EnKF which results in better performance than the projection method. This method is similar to the constrained method used in the UKF. The methods introduced in this chapter are demonstrated on a continuous-discrete CSTR process.

3.1 Introduction

In this chapter, we consider another nonlinear filtering algorithm known as the ensemble Kalman filter (EnKF). The EnKF, which was first introduced by Evensen [9] in 1994, uses Monte Carlo sampling method when generating the initial ensemble, the process noise ensemble and the measurement noise ensemble. The ensemble of sample points are then propagated through the nonlinear system and the probability distribution of the state is recovered from the samples. Table 3.1 briefly summarizes some of the common filters in the literature.

Table 3.1: Classification of filters

	Kalman	Bayes' rule
No sampling	EKF	
Deterministic sampling	UKF Resample at each iteration	
Monte Carlo sampling	EnKF Draw samples only once	Particle filter

In contrast with the EKF, the EnKF represents the error covariance matrix by an ensemble of samples. Thus, the uncertainty is represented by a set of model realizations in the EnKF, rather than an explicit expression for the error covariance matrix as in the EKF. The EKF approximation can introduce large errors in the covariance because of the model nonlinearity. In the EnKF, the approximation of forecasting the mean and covariance is accomplished by the propagation of the ensemble of samples forward in time.

Both the EnKF and the UKF belong to a broader category of filters called derivative-free filters. In derivative-free filters, better estimates of the moments of a distribution can be obtained using samples rather than using the Taylor series approximation of the nonlinear function. The main difference between the EnKF and the UKF is that the former uses stochastic sampling method while the latter uses a deterministic sampling method. A large number of samples are necessary for generating good estimates in the EnKF. Moreover, the EnKF draws samples only once at the initialization, and the state estimate and covariance can be recovered from the corresponding samples while running the filter.

The EnKF is widely used in applications like weather forecasting, where the models are of extremely high order and nonlinear, the initial states are highly uncertain, and a large number of measurements are available [10]. The EnKF has been shown to be very efficient and robust for real-time updating in weather forecasting [28], oceanography [29] and meteorology [30]. Also, in petroleum reservoir simulation, Nævdal et al. [31] implemented the EnKF in monitoring of the near-well zones in an oil reservoir, in order to estimate the reservoir's permeability distribution. The success indicates that the EnKF is capable of handling highly nonlinear and complex systems [32].

The conventional solution for constraint incorporation in estimators such as the Kalman filter, the EKF and the UKF is the use of the projection method. In the projection method, the estimates which violate the constraints are projected onto the constraint boundaries. However, for the EnKF, where an ensemble of particles are involved, an appropriate projection method which provides convergence in the estimation performance must be used. Besides the projection method, moving horizon estimation (MHE) can naturally take constraints into consideration, but the computational load remains an issue for high-dimensional systems. Moreover, accurate arrival cost approximation is difficult in the presence of constraints.

The main contributions of this chapter are as follows: We propose an appropriately modified projection method for the samples in the EnKF that retains the unconstrained covariance, by which the EnKF can provide the convergence in the constrained estimation. We then propose an alternative constrained parameter estimation method with the EnKF. This method exhibits better performance than the projection method. The efficacies of these methods are demonstrated on a simulated chemical process.

This chapter is organized as follows. First, in Section 3.2, we introduce three different ways, namely the mean and covariance, the sigma points, the Monte Carlo samples, for representing the distribution of an estimate. In Section 3.3, the EnKF algorithm is presented and its performance for parameter estimation is assessed. The influence of the ensemble size for the EnKF is studied, and it is also compared with the UKF. In Section 3.4, constrained estimation with the EnKF is considered.

We propose a modified projection method that works with the EnKF. Then, a new constrained algorithm is proposed which provides better performance than the projection method when applied to the EnKF. Finally, a continuous-discrete CSTR process is employed to demonstrate the performance of these methods.

3.2 Preliminaries

3.2.1 Estimate representation

In state estimation, we normally use the first two moments, the mean vector x_k and the covariance matrix P_k, to represent an underlying distribution. An accurate covariance is critical to the performance of an estimator. A simple illustration of the mean and covariance is shown in Fig.3.1.

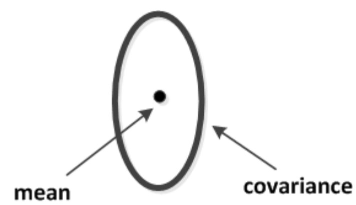


Figure 3.1: Illustration of mean and covariance of a two dimensional estimate

In the unscented Kalman filter (UKF), a set of deterministic samples, called sigma points, are employed to represent the distribution of an estimate. We should note that both the number of the sigma points, the values of the sigma points, as well as the weights of the sigma points are calculated based on a deterministic method. For instance, five weighted sigma points should be selected for a two dimensional distribution. Details of the selection method and weights calculation can be found in the Chapter 2. An illustration of using sigma points to represent the distribution is shown in Fig.3.2.

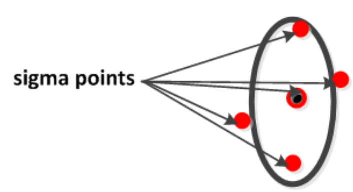


Figure 3.2: Sigma points representation of a two dimensional estimate

Besides the sigma point representation, we can use the Monte Carlo approach to build the necessary statistics. In Monte Carlo sampling method, an ensemble of particles is obtained by randomly drawing from a distribution. The number of the particles are chosen according to demands of accuracy, and each of the particles is equally weighted. Fig.3.3 is an illustration of the Monte Carlo sampling method.

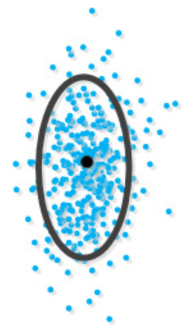


Figure 3.3: Monte Carlo samples of a two dimensional estimate

Once we have these particles, we can recover the estimate and its covariance by calculating the ensemble mean and covariance. For instance, say have an ensemble X of size N,

$$X = [x^1, \dots, x^N] \tag{3.1}$$

where X is an $n \times N$ matrix (n is the dimension of the state). The recovered mean and covariance can be calculated by Eq.3.2 and Eq.3.3, respectively.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x^i \tag{3.2}$$

$$P = \frac{1}{N-1}\sum_{i=1}^{N}(x^i - \bar{x})(x^i - \bar{x})^T \tag{3.3}$$

It is essential that the ensemble is statistically representative. Suppose now that we have a normal distribution with mean 0 and variance 1. The Monte Carlo samples with ensemble size 100 and ensemble size 10000 are shown in Fig.3.4. It is seen that the distribution can be better recovered with larger ensemble size. When the ensemble size N increases, the estimate representation error will decrease proportional to $1/\sqrt{N}$.

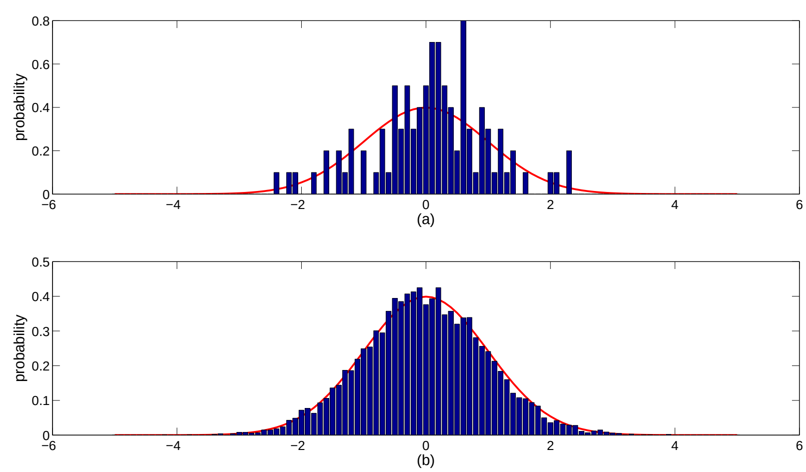


Figure 3.4: Monte Carlo sampling comparison with different ensemble sizes ((a) 100, (b) 10000)

3.2.2 Ensemble Kalman filter

The ensemble Kalman filter (EnKF) belongs to a broader category of filters known as derivative-free filters. The EnKF is initialized by choosing a number of random particles that capture the initial probability distribution, as described above. These particles are propagated through the nonlinear model, and the estimates (prior and posterior) are approximated by the ensemble mean and covariance.

Let us again consider the following nonlinear model,

$$x_{k+1} = f(x_k, u_k) + w_k \tag{3.4a}$$
$$y_k = h(x_k) + v_k \tag{3.4b}$$

where w_k and v_k are uncorrelated zero-mean white noise with covariance matrices Q and R, respectively.

The filter is initialized by generating an initial ensemble. In this work, we use `mvnrnd` in MATLAB to draw the initial ensemble from an initial guess of the estimate (x_0 and P_0). Implementation of the EnKF requires an adequate number of particles, since a small number of particles may not guarantee a good approximation of the true covariance matrix. Where the ensemble size is too small to be statistically representative of the estimate, it is said to be underestimated. Underestimation is a fundamental problem in the EnKF [33]. On the other hand, a large number of particles require extensive storage and computing resources. Once we have obtained the initial ensemble, the EnKF algorithm can be applied.

The EnKF consists of three steps:

1. a prediction step that propagates the ensemble cloud through the model to generate a prior ensemble

2. an analysis step that computes the Kalman gain based on the ensemble statistics

3. an update step that updates the prior ensemble to the posterior ensemble with the measurement information

Prediction

In the prediction step, the particles $[x_k^1, \ldots, x_k^N]$ are propagated one step forward through the state model f. Also, we consider that the nonlinear model is not perfect and contains model errors, which is represented by w_k. In order to fully capture the propagation of the state particles, an ensemble of particles of w_k is also generated based on its distribution $\mathcal{N}(0, Q)$. The state noise particles are denoted by $[w_k^1, \ldots, w_k^N]$.

Thus, the predicted state ensemble is calculated by

$$x_{k+1}^{i-} = f(x_k^i, u_k) + w_k^i \tag{3.5}$$

Also, the predicted output is calculated based on the measurement model

$$y_{k+1}^{i-} = h(x_{k+1}^{i-}) \tag{3.6}$$

As we have the predicted state ensemble from Eq.3.5, the prior state estimate x^-_{k+1} and prior covariance P^-_{k+1} can be obtained from Eq.3.7 and Eq.3.8, respectively.

$$x^-_{k+1} = \frac{1}{N}\sum_{i=1}^{N} x^{i-}_{k+1} \tag{3.7}$$

$$P^-_{k+1} = \frac{1}{N-1}\sum_{i=1}^{N}(x^{i-}_{k+1} - x^-_{k+1})(x^{i-}_{k+1} - x^-_{k+1})^T \tag{3.8}$$

In the prediction step, an ensemble of prior state particles and predicted outputs are generated. The prior state estimate (mean and covariance) can be recovered from the prior state particles.

Analysis

In the analysis step, we can obtain the Kalman gain using the ensemble statistics. The output error covariance matrix is defined as the sample covariance of y^{i-}_{k+1} around the its sample mean. The sample mean of the output ensemble is calculated by

$$y^-_{k+1} = \frac{1}{N}\sum_{i=1}^{N} y^{i-}_{k+1} \tag{3.9}$$

The output error covariance P^{yy} is calculated by

$$P^{yy}_{k+1} = \frac{1}{N-1}\sum_{i=1}^{N}(y^{i-}_{k+1} - y^-_{k+1})(y^{i-}_{k+1} - y^-_{k+1})^T \tag{3.10}$$

Similarly, we need to calculate the cross covariance matrix P^{xy} as

$$P^{xy}_{k+1} = \frac{1}{N-1}\sum_{i=1}^{N}(x^{i-}_{k+1} - x^-_{k+1})(y^{i-}_{k+1} - y^-_{k+1})^T \tag{3.11}$$

Both the output error covariance P^{yy} and cross-covariance P^{xy} have similar meaning as in the unscented Kalman filter. In the EnKF, they are calculated based on the equally weighted particles rather than the weighted sigma points used in the UKF. The analysis step of the EnKF calculates the Kalman gain K as follows:

$$K = P^{xy}_{k+1}(P^{yy}_{k+1} + R)^{-1} \tag{3.12}$$

The above form of the Kalman gain is a modification of the standard Kalman gain represented as $K = P^-H^T(HP^-H^T + R)^{-1}$. One of the advantages of the EnKF is that the prior covariance P^- is not required in the calculation of K.

Update

In the update step, each of the particles in the predicted ensemble is updated independently with the information of the incoming measurement y_{k+1}. In order to update the particles properly, it is critical to perturb the measurement

y_{k+1}. It has been shown that unless a measurement ensemble is generated at each iteration, the updated ensemble will have a covariance that is too low, which can result in a large bias [34]. The perturbed measurement ensemble is calculated by

$$y_{k+1}^i = y_{k+1} + v_k^i \tag{3.13}$$

where v_k^i represents the measurement noise for the ith particle, which follows $\mathcal{N}(0, R)$.

Therefore, each particle can be updated using the following equation:

$$x_{k+1}^i = x_{k+1}^{i-} + K(y_{k+1}^i - y_{k+1}^{i-}) \tag{3.14}$$

Similarly, the posterior mean and covariance can be recovered from Eq.3.15 and Eq.3.16 respectively.

$$x_{k+1} = \frac{1}{N} \sum_{i=1}^{N} x_{k+1}^i \tag{3.15}$$

$$P_{k+1} = \frac{1}{N-1} \sum_{i=1}^{N} (x_{k+1}^i - x_{k+1})(x_{k+1}^i - x_{k+1})^T \tag{3.16}$$

Eq.3.5 to Eq.3.16 above represent the ensemble Kalman filtering algorithm, and a schematic is provided in Fig.3.5.

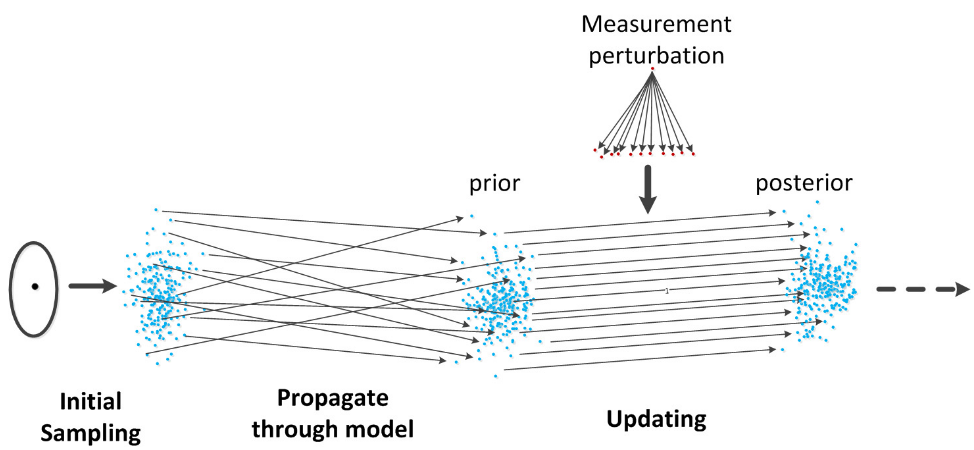


Figure 3.5: Schematic of the ensemble Kalman filter (EnKF)

It should be noted that errors in the EnKF come from two sources, both related to the sampling of the probability distribution: 1. initial representation of the distribution by particles; 2. state and measurement noise representation by particles. A large ensemble size will be helpful in reducing the errors in the EnKF. The EnKF will have identical performance to the Kalman filter if the ensemble size goes to infinity for a linear system.

Both the UKF and the EnKF belong to the class of derivative-free filters, but major differences between the UKF and the EnKF lie in

1. The method of drawing samples

2. The choice of re-drawing samples
3. The Kalman gain calculation
4. The need for a perturbed measurement ensemble
5. The method used for ensemble update from prior to posterior

Moreover, the prior and posterior mean x^- and x, and prior and posterior covariance P^- and P are not necessarily to obtain in the EnKF in each iteration. All of them can be recovered by the corresponding ensemble.

3.3 EnKF for parameter estimation

Parameter estimation, or dual state and parameter estimation, has evoked significant interest in process control. Generally, parameter estimation is a nonlinear estimation problem solved by augmenting the unknown parameters as additional model states. Typically, random walk models are used to describe the dynamics/uncertainties of the parameters of interest. We have shown earlier that the extended Kalman filter (EKF) and unscented Kalman filter (UKF) are powerful tools to estimate states and parameters simultaneously. In this section, a simulation of a continuous stirred tank reactor (CSTR) with discrete measurements is employed to assess the parameter estimation performance of the EnKF.

The model of the CSTR process is given by [26]

$$\dot{C}_A = \frac{q}{V}(C_{Af} - C_A) - k_0 e^{-\frac{E}{RT}} C_A + w_1 \tag{3.17a}$$

$$\dot{T} = \frac{q}{V}(T_f - T) + \frac{(-\Delta H)}{\rho C_p} k_0 e^{-\frac{E}{RT}} C_A + \frac{UAr}{V\rho C_p}(T_c - T) + w_2 \tag{3.17b}$$

This process has two states, concentration C_A and temperature T. ρ and $\frac{E}{R}$ are assumed to be unknown parameters that are to be estimated. The random walk models for them are

$$\dot{\rho} = 0 + w_3 \tag{3.18}$$

$$\frac{\dot{E}}{R} = 0 + w_4 \tag{3.19}$$

Two discrete measurements are available as

$$y_{1,k} = C_{A,k} + v_{1,k} \tag{3.20}$$

$$y_{2,k} = T_k + v_{2,k} \tag{3.21}$$

The initial guesses for ρ and $\frac{E}{R}$ are 1025 and 8755, while their true values are 1000 and 8750, respectively. Both the unknown parameters are treated as augmented states. Hence, in the parameter estimation of this CSTR process, we have four states to be estimated. Covariance matrices for the process and measurement zero-mean Gaussian noise are set as

$$Q = \begin{bmatrix} 1\times10^{-5} & 0 & 0 & 0 \\ 0 & 1\times10^{-3} & 0 & 0 \\ 0 & 0 & 1\times10^{-10} & 0 \\ 0 & 0 & 0 & 1\times10^{-10} \end{bmatrix}, \qquad R = \begin{bmatrix} 1\times10^{-4} & 0 \\ 0 & 1\times10^{-2} \end{bmatrix}$$

We choose different ensemble sizes to evaluate the performance of the EnKF for parameter estimation with the same initial condition. The first EnKF uses 100 particles to represent the distribution, while the second EnKF uses 400 particles. The simulation results are shown in Fig.3.6 and Fig.3.7 for parameters ρ and $\frac{E}{R}$ respectively. A comparison with the performance of the UKF is also provided.

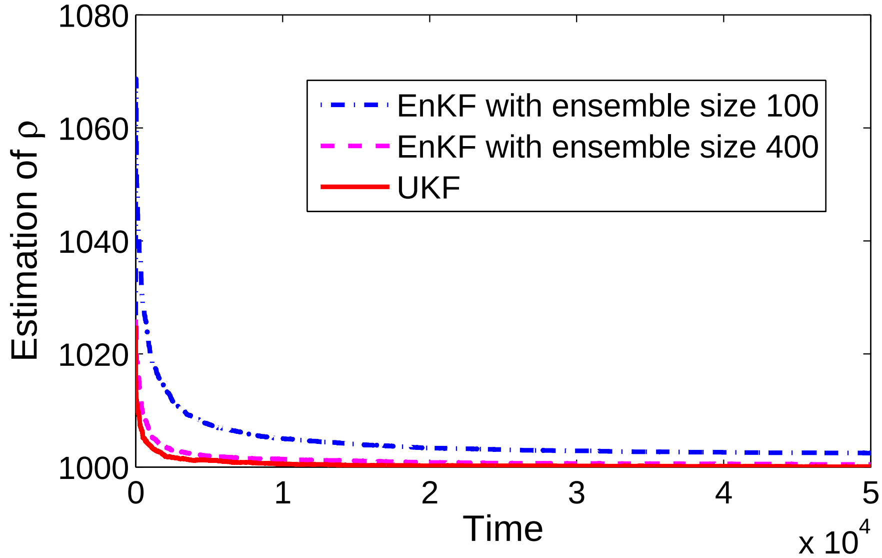


Figure 3.6: Comparison of the estimation performance for ρ in the CSTR model

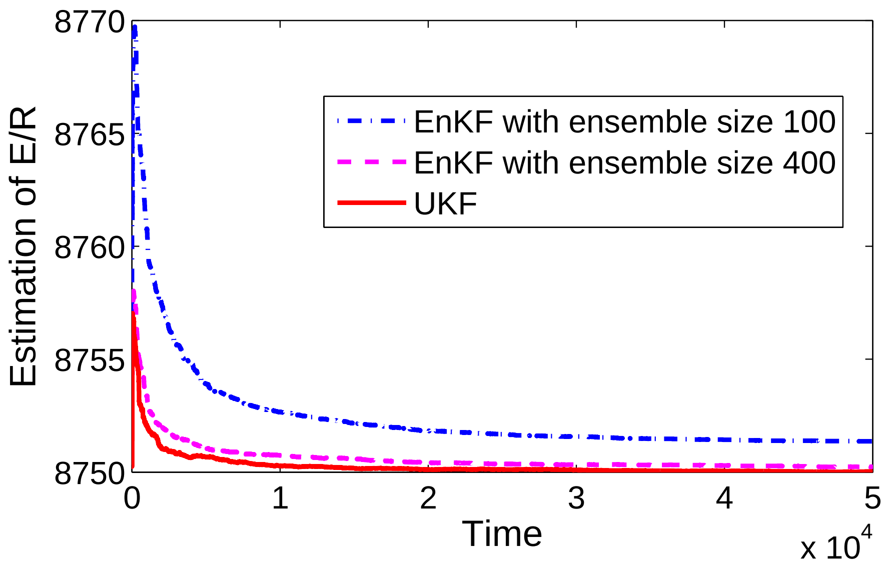


Figure 3.7: Comparison of the estimation performance for $\frac{E}{R}$ in the CSTR model

Fig.3.6 and Fig.3.7 show that the larger the ensemble size we employ, the better the performance of the EnKF has. The UKF, which uses weighted deterministic sigma points rather than equally weighted random points, can achieve similar performance with less computational load.

3.4 Inequality constrained parameter estimation with the EnKF

In practical applications, parameter constraints are commonly encountered. In this work, we consider inequality parameter constraints shown as

$$d^L \leq c(\theta) \leq d^U \tag{3.22}$$

where d^L and d^U indicate the lower and upper bound of the inequality constraints. Implementation of such parameter constraints in estimation will be useful for improving estimation performance. We have introduced a method to construct inequality parameter constraints in Chapter 2. In this section, we will discuss several ways of implementation of inequality constraints in the EnKF, and propose a proper projection method and a constrained parameter estimation method with the EnKF.

3.4.1 Constrained EnKF algorithm

Prakash et al. [21] proposed a constrained EnKF algorithm for state estimation. The approach includes using a truncated multivariate distribution in the presence of state constraints. Furthermore, both the projection method and the optimization-based solution are employed in this algorithm to solve the constrained nonlinear state estimation problem.

In this method, the ensemble particles will propagate through the nonlinear model f to obtain the prior ensemble. Then, the propagated particles which lie outside the feasible region are projected onto the constraint boundaries to obtain the constrained prior ensemble. This is shown in Eq.3.23

$$x_c^{i-} = Pr[x^{i-}] \tag{3.23}$$

Then, in the update step, the ensemble of constrained state estimates (posterior) is obtained by solving a number of constrained optimization problems. The optimization problem for each particle is shown below:

$$x^i = argmin_{x^i}(y - h(x^i))^T R^{-1}(y - h(x^i)) + (x^i - x_c^{i-})^T P^{-1}(x^i - x_c^{i-}) \tag{3.24}$$

subject to the constraints.

The efficacy of this method for constrained nonlinear state estimation has been shown in [21]. However, for the constrained parameter estimation problem, the parameter estimates may not be properly updated using Eq.2.53 once the parameter constraints are active. This method fails mainly because the first term in the objective function is not even needed, because y is only a function of x. Therefore, this method may result in infeasible or inaccurate estimates for parameter estimation. Simulation studies will demonstrate its failure in the constrained parameter estimation.

3.4.2 Projection method

The projection (or clipping) method is very common in accounting for constraints. In this section, we will investigate its possible use in the EnKF.

In the EnKF, a number of particles are involved in both the prediction and the update steps. The unconstrained particles are given as

$$X = [x^1, x^2, \ldots, x^N] \tag{3.25}$$

We use

$$\tilde{X} = [\tilde{x}^1, \tilde{x}^2, \ldots, \tilde{x}^N] \tag{3.26}$$

to denote the constrained particles.

In most filtering algorithms, the covariance plays an important role in properly updating the estimate. Some constraint implementation methods may influence the covariance significantly and lead to poor performance of the estimator. Below, we will show three different methods of projection in the EnKF, and evaluate them in the sense of covariance consistency.

First method Projection of all the particles

The projection method projects the violated particles onto the constraints boundaries. In this way, the covariance of the constrained ensemble will be different from that of the unconstrained ensemble. We use the ensemble of parameter ρ to illustrate the performance of this projection method. The constrained particles are subject to the specified nonlinear inequality constraints shown in Section 2.5.1. The trajectories of unconstrained particles and constrained particles of ρ are shown in Fig.3.8 (both of them have ensemble sizes equal to 400). We also have the unconstrained distribution of the ensemble in Fig.3.9(a) and the constrained distribution in Fig.3.9(b). It is seen that the value of the ensemble is changed due to the constraint, and the distribution is significantly changed, too. The variance of the constrained particles, in this case, is less than that of the unconstrained particles. This kind of distribution change can lead to the inaccuracy of estimation result.

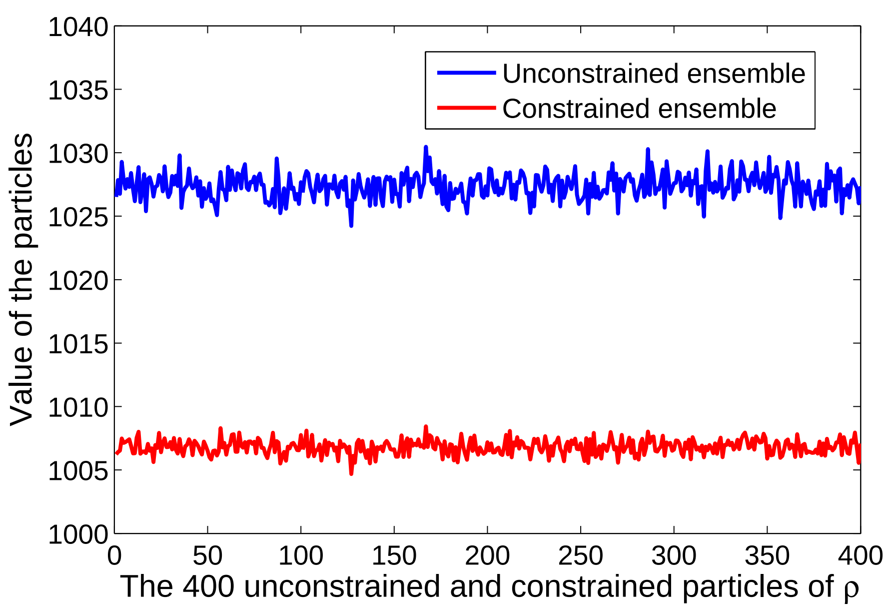


Figure 3.8: Plot of unconstrained and constrained particles of ρ

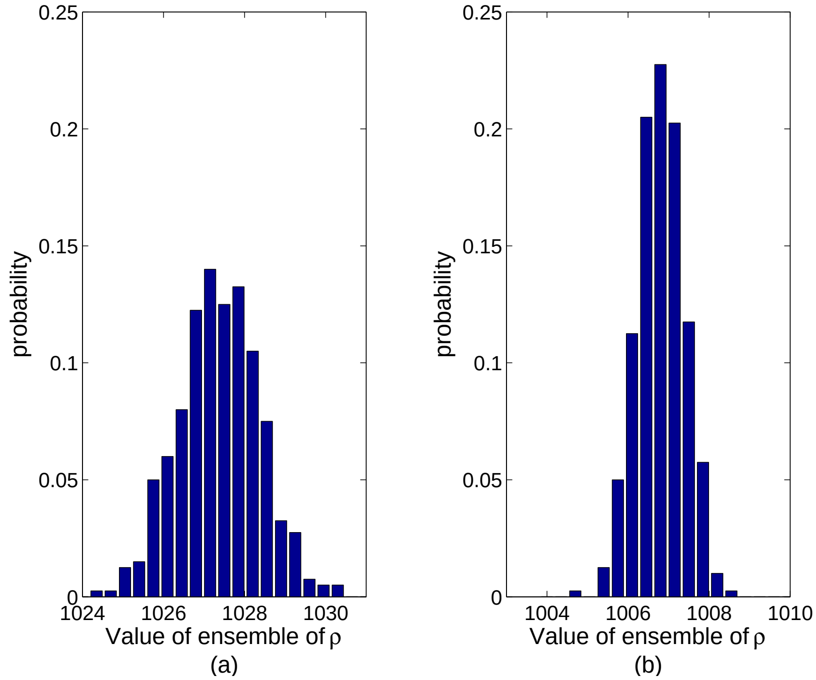


Figure 3.9: Distribution of the ensemble of parameter ρ before and after applying constraints by the first projection method

Second method Projection of the mean value and redrawing particles

In order to avoid significant change in the covariance while imposing the constraints, one can simply constrain the estimated mean value of the particles onto the constraint boundaries, and then use the same covariance to redraw particles. In this method, there will not be much change in the covariance, but we cannot guarantee that all the particles fall into the constrained space. The illustration of this method is shown in Fig.3.10. Fig.3.10(a) shows the unconstrained distribution of the ensemble while the constrained distribution of the ensemble is shown in Fig.3.10(b). It is seen that the covariances remain similar in this method, although they are not identical. This can lead to a bias in the estimation result, and a simulation will be shown later to confirm this.

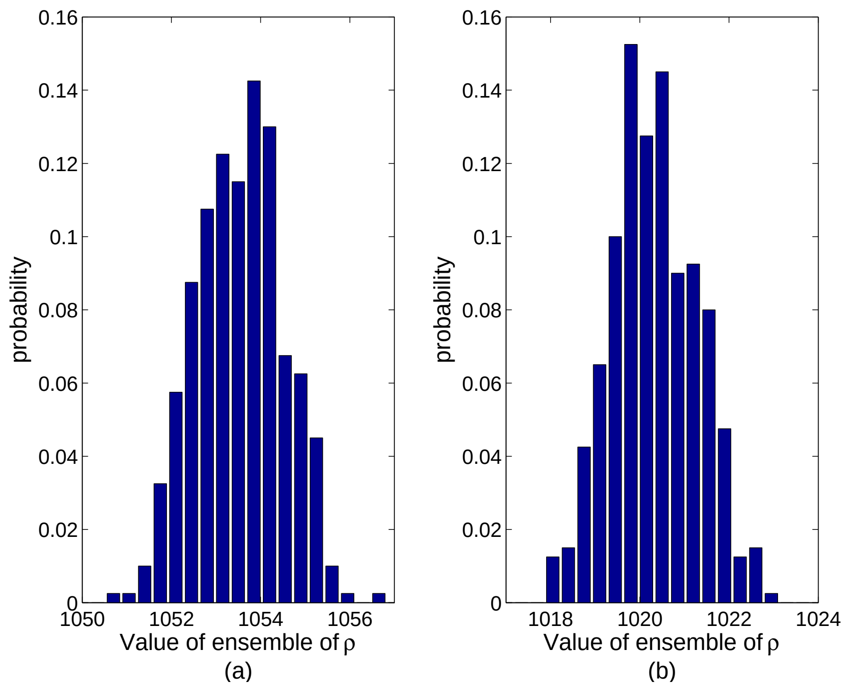


Figure 3.10: Distribution of the ensemble of parameter ρ before and after applying constraints by the second projection method

Third method Projection of the mean value and shifting all the particles by the same difference

It has been discussed above that a change of covariance may result in poor estimation performance; therefore, we need to find a way to obtain the identical covariance after constraint incorporation using the projection method. In this method, we also constrain the mean value of the unconstrained particles onto the constraints boundaries. Then, we calculate the difference D as

$$D = \tilde{x} - \bar{x} \tag{3.27}$$

where $\bar{x}$ is the mean value of unconstrained particles, and $\tilde{x}$ is the constrained mean value. Once we obtain the difference D, we can obtain the shifted particles by

$$\tilde{x}^i = x^i + D \tag{3.28}$$

In this method, the mean value of the constrained particles falls on the constraints boundaries and the covariance after imposing the constraint remains the same. This method also cannot guarantee that all the particles fall into the constrained space. This method is illustrated in Fig.3.11, where Fig.3.11(a) represents the unconstrained distribution of the ensemble while the constrained distribution of the ensemble is shown in Fig.3.11(b). It is seen that the values of particles are changed due to the constraints, while the covariances are identical. This method can provide improved constrained parameter estimation performance compared to the second projection method.

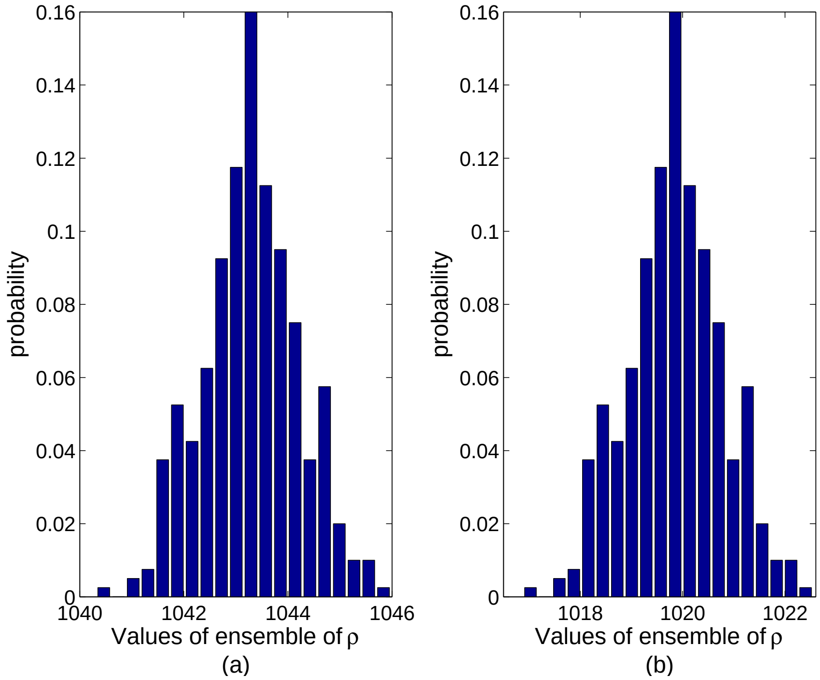


Figure 3.11: Distribution of the ensemble of parameter ρ before and after applying constraints by the third projection method

3.4.3 Proposed method

Although the third method presented in the previous section is the best choice of the three methods discussed, it is based on simple projection. In this section, a further improved method will be proposed. This method is similar to the constrained UKF method proposed in the previous chapter. In this method, the mean value of the constrained parameter estimates will fall into the constrained space, rather than on the constraint boundaries as in the projection method. The constrained mean value is obtained by solving

$$\tilde{x}_k = argmin_{x_k}(y_k - \hat{y}_k)^T R^{-1}(y_k - \hat{y}_k) + (x_k - \bar{x}_k)^T I(x_k - \hat{x}_k) \tag{3.29}$$

subject to the inequality parameter constraints

$$d^L \leq c(\theta) \leq d^U \tag{3.30}$$

where $\hat{y}_k$ is the estimated output from the previous step $k-1$ with decision variable θ. By solving the optimization problem, a parameter θ which falls into the constrained space can be obtained. In other words, the mean value of the constrained particles is more precisely recovered than in the third projection method as stated in the previous section. Then, we shift all the particles by the difference between the constrained mean value and the unconstrained mean value, to get an identical covariance. A simulation result will be presented later to demonstrate the superiority of this method.

3.5 Simulation example

In this section, we consider the inequality constrained parameter estimation of the CSTR model. The methods described in Section 3.4 will be applied to demonstrate their performance. The inequality parameter constraints come from the steady-state measurements of the process and are shown below:

$$\begin{bmatrix} 0.9451 \\ 312.6003 \end{bmatrix} \leq \begin{bmatrix} c_1(\theta) \\ c_2(\theta) \end{bmatrix} \leq \begin{bmatrix} 0.9570 \\ 312.7239 \end{bmatrix} \tag{3.31}$$

Detail of the constraint generation can be found in Chapter 2.

First, we apply the constrained EnKF method proposed by Prakash et al. [21]. This method includes projection of the prior ensemble and updating of the ensemble by solving a set of optimization problems. The objective function of the optimization problem is not feasible for constrained parameter estimation. Fig.3.12 and Fig.3.13 are the simulation results of estimating parameters ρ and $\frac{E}{R}$ with this constrained EnKF method. It is obvious that this method leads to large inaccuracy in constrained parameter estimation.

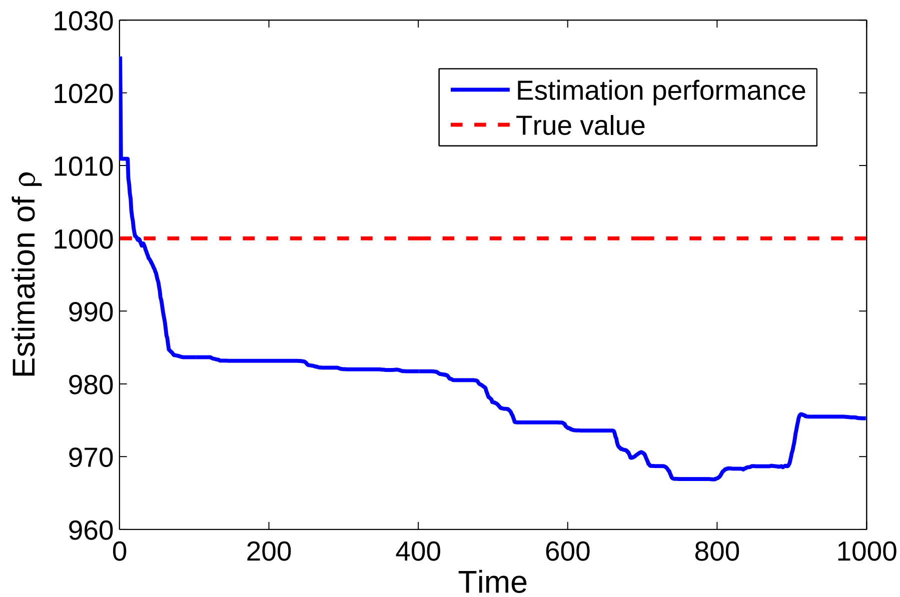


Figure 3.12: Inequality constrained estimation of parameter ρ in the CSTR model by the constrained EnKF proposed by Prakash et al.

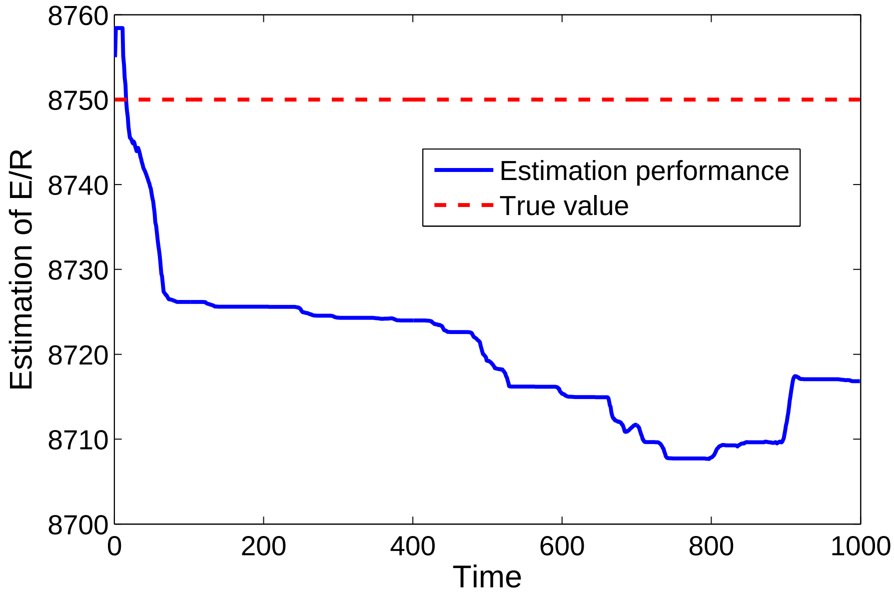


Figure 3.13: Inequality constrained estimation of parameter $\frac{E}{R}$ in the CSTR model by the constrained EnKF proposed by Prakash et al.

Then, the first projection method is applied. In this method, after obtaining the unconstrained ensemble with the EnKF at each sampling instant, the constrained ensemble is calculated based on Eq.4.25. Also, the second projection method is applied. In this method, the ensemble mean is constrained and the ensemble particles are re-drawn with the same covariance. Simulation results are shown in Fig.3.14 and Fig.3.15 for parameters ρ and $\frac{E}{R}$ respectively. It is seen that bias is present in the estimation performance.

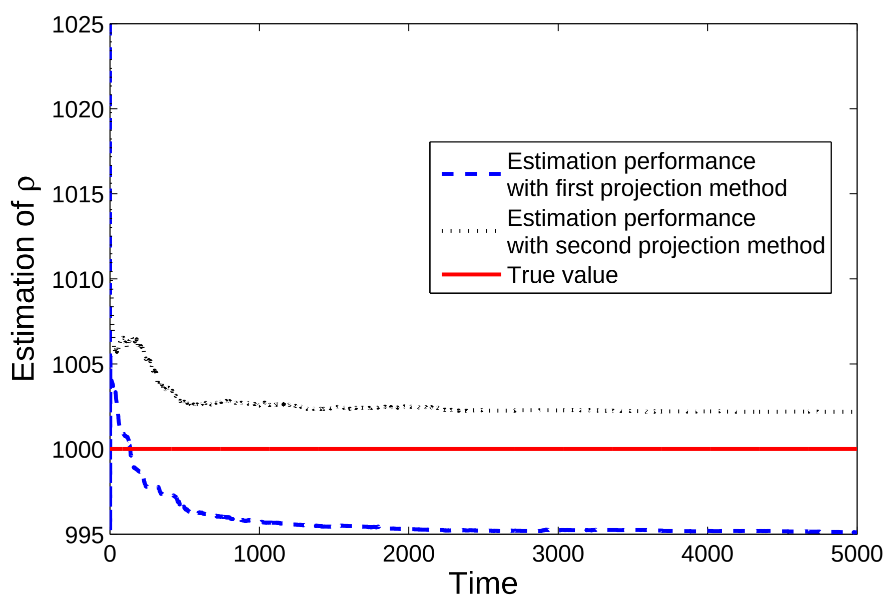


Figure 3.14: Inequality constrained estimation of parameter ρ in the CSTR model by the first and second projection methods

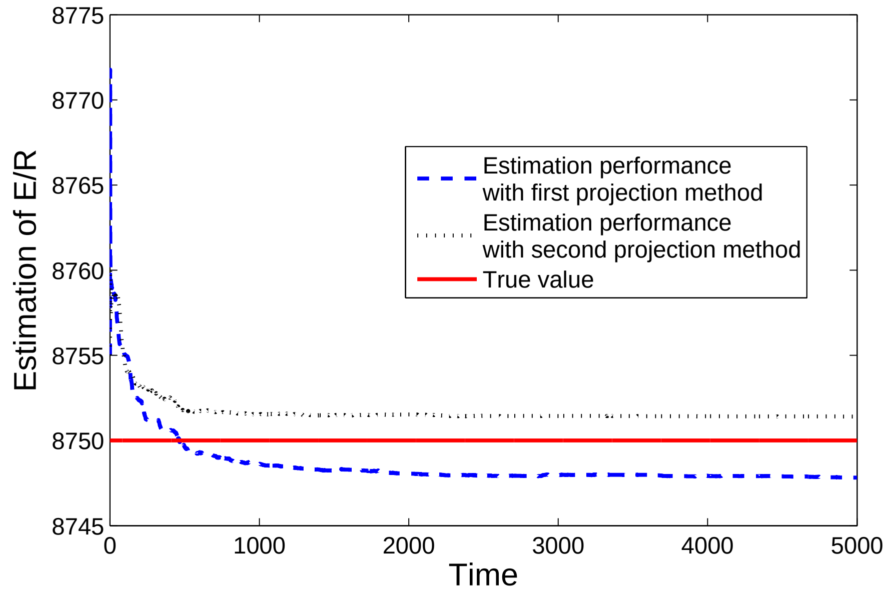


Figure 3.15: Inequality constrained estimation of parameter $\frac{E}{R}$ in the CSTR model by the first and second projection methods

Next, we use the third projection method to estimate the inequality constrained parameters. In this method, the mean of the estimate is constrained onto the constraints boundaries, while an identical covariance is preserved for the ensemble. Fig.3.16 and Fig.3.17 are the simulation results of estimation of parameters ρ and $\frac{E}{R}$. This projection method provides the convergence of the estimation.

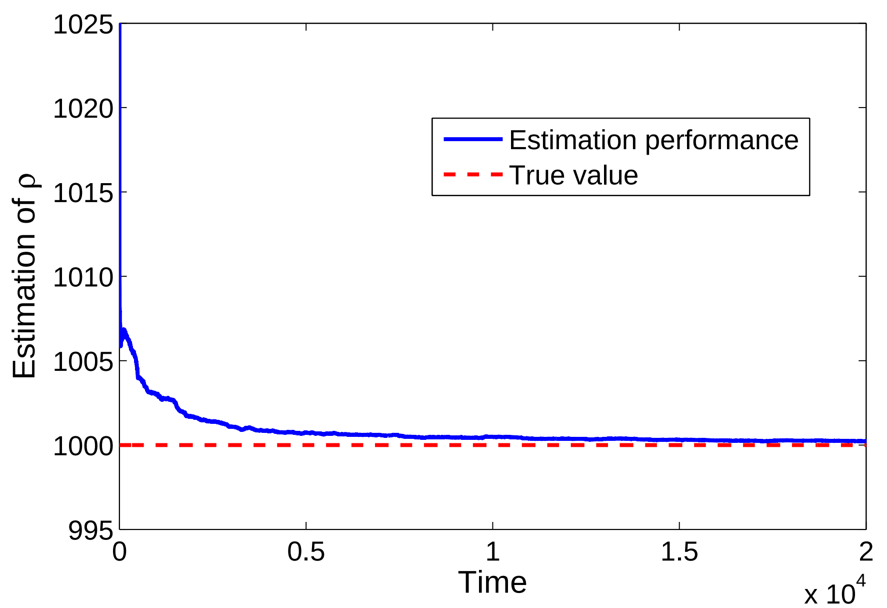


Figure 3.16: Inequality constrained estimation of parameter ρ in the CSTR model by the third projection method

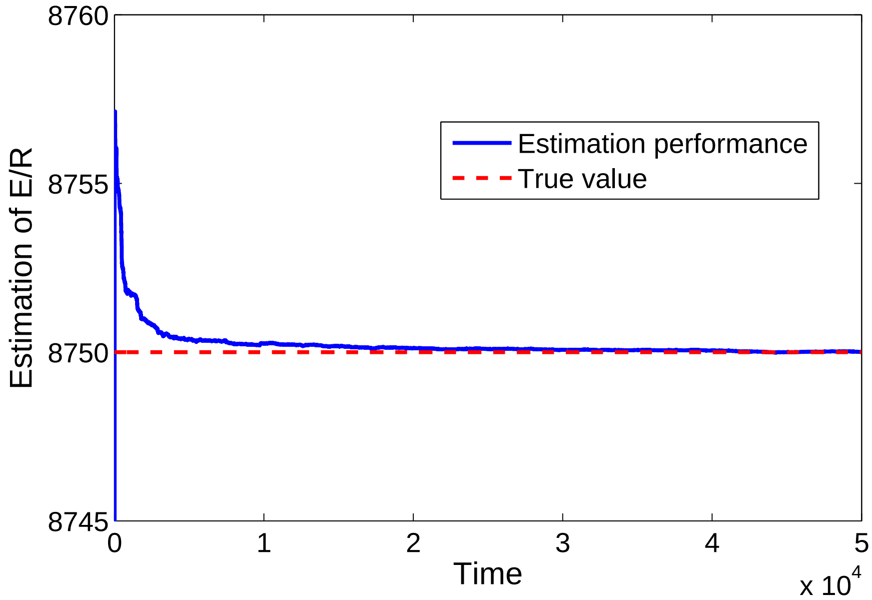


Figure 3.17: Inequality constrained estimation of parameter $\frac{E}{R}$ in the CSTR model by the third projection method

Finally, the proposed method is applied, and it results in faster recovery of estimates compared to the projection method. In the proposed method, the constrained parameter estimates can fall into the constrained space, rather than on the constraints boundaries as in the projection method. Fig.3.18 and Fig.3.19 are simulation results compared to the third projection method. It is shown that the proposed method can improve the performance of the constrained parameter estimation.

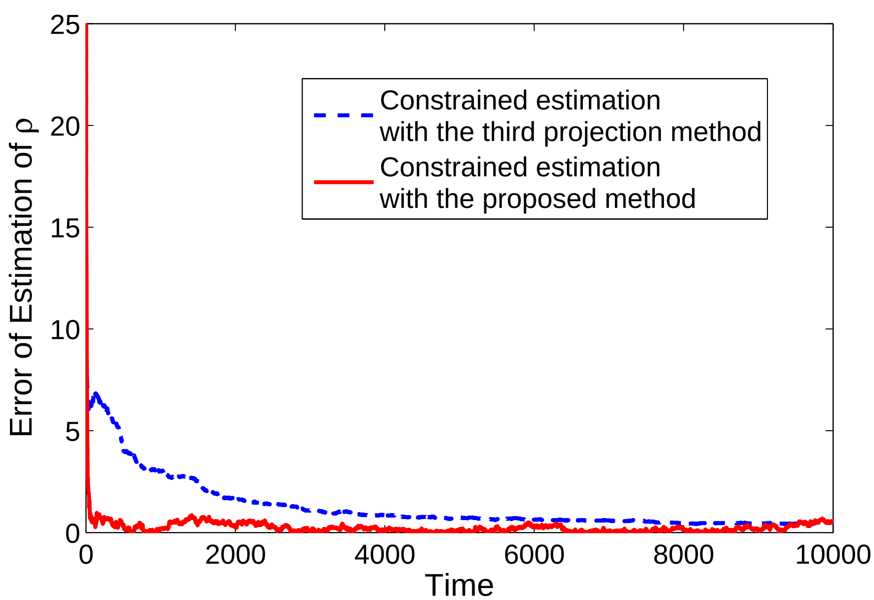


Figure 3.18: Inequality constrained estimation of parameter ρ in the CSTR model by the proposed method

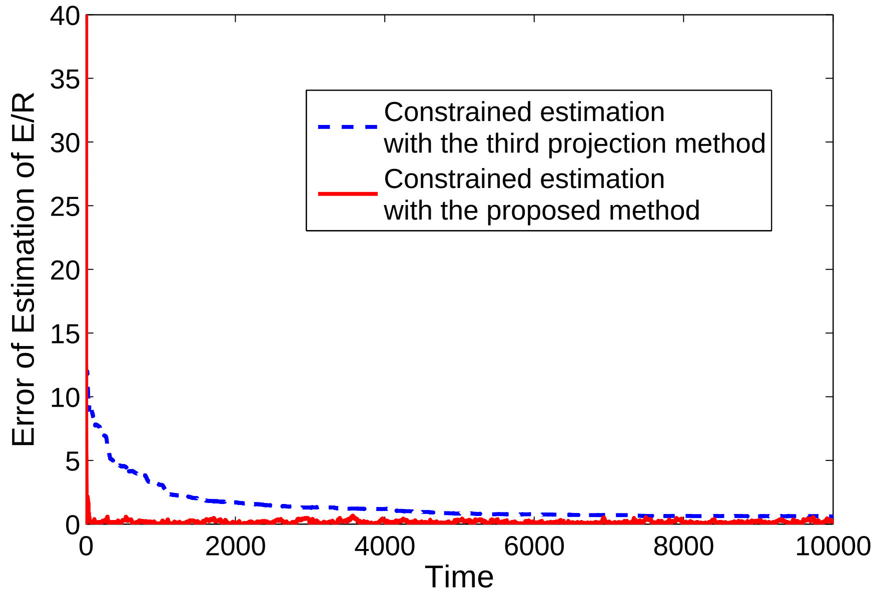


Figure 3.19: Inequality constrained estimation of parameter $\frac{E}{R}$ in the CSTR model by the proposed method

3.6 Conclusions

Constrained parameter estimation with the ensemble Kalman filter (EnKF) has been considered in this chapter. A more appropriate use of the projection method in constraining the particles in the EnKF is introduced. In constraints incorporation, an identical covariance should be generated in order to obtain better convergence in the estimation result. An inequality constrained parameter estimation method is proposed. The proposed method can provide faster recovery of parameter estimates to account for parameter constraints compared to the projection method. The introduced projection method and the proposed method are compared in the estimation performance of a continuous-discrete CSTR process, where the performance of the proposed method is shown to be superior.

Chapter 4

Inequality Constrained Parameter Estimation with Moving Horizon Estimation

Moving horizon estimation (MHE) for constrained parameter estimation is considered in this chapter. We first present the MHE algorithm, as well as common approaches for arrival cost approximation. We then address the problem with MHE when incorporating constraints directly. The proposed method provides an alternative way for constrained parameter estimation with MHE. The estimation performance is demonstrated on a continuous-discrete CSTR model.

4.1 Introduction

In this chapter, we consider another estimation method known as moving horizon estimation (MHE). MHE is an optimization based strategy for state estimation. MHE can be regarded as the dual of model predictive control (MPC). The success of employing on-line optimization in industrial MPC provided the motivation for MHE. Unconstrained state estimation with moving horizon was proposed by Thomas [35], and later on by Kwon, Bruckstein and Kailath [36]. Moving horizon state estimation was first applied in nonlinear initial state estimation in a noise free environment without constraints by Jang, Joseph and Mukai in 1986 [37].

Solving the optimization problem, however, is computationally demanding, because the problem dimension grows with time as more data are processed. One method to reduce the computational load is to bound the size of the estimation problem by employing a moving horizon approximation. In moving horizon estimation, the state estimate is determined on-line by solving a finite horizon optimization problem. As new measurements become available, the old measurements are discarded from the estimation window, and the finite horizon state estimation problem is solved again to determine the new estimate of the state.

Because MHE is formulated as an optimization problem, it is possible to explicitly handle inequality constraints. Muske and Rawlings [38] derived some preliminary conditions for the stability of state estimation with inequality constraints. Tyler and Morari [39] demonstrated how constraints may result in instability for nonminimum phase systems. For the constrained problem, unfortunately, it is not possible to generate an analytic expression for the arrival cost. In addition, it is difficult to include the information of constraints into the arrival cost.

The rest of this chapter is organized as follows. First, we introduce the problem formulation and the moving horizon estimation (MHE) algorithm in Section 4.2. In Section 4.3, various arrival cost approximations in unconstrained MHE are investigated. Then, the constrained estimation with MHE is considered in Section 4.4. The problem of directly constrained MHE lies in the approximation of the arrival cost, and it will introduce large bias in the estimation result. We propose an alternative method to perform the constrained parameter estimation with MHE, which provides better performance than the directly constrained MHE. Simulation results illustrating their performance are shown in Section 4.5.

4.2 Moving horizon estimation

4.2.1 Problem formulation

The estimation problem is considered again in this section. The process dynamics are given by the following equations

$$x_{k+1} = f(x_k, u_k) + w_k \tag{4.1a}$$
$$y_k = h(x_k) + v_k \tag{4.1b}$$

where

$$w_k \sim \mathcal{N}(0, Q) \tag{4.2}$$
$$v_k \sim \mathcal{N}(0, R) \tag{4.3}$$

The state function, f, and the measurement function, h, are not necessarily linear. The estimate of the state x_k given measurements $\{y_1, y_2, \ldots, y_k\}$ can be achieved using different kinds of algorithms. The Kalman filter can provide optimal estimation performance if we have linear functions f and h. For nonlinear estimation problems, the extended Kalman filter (EKF), the unscented Kalman filter (UKF) or the ensemble Kalman filter (EnKF) can be applied to obtain sub-optimal estimation solutions. The EKF, the UKF and the EnKF are based on the prediction-update Kalman structure. Their solutions are not optimal because there will be error when representing the real state distribution.

Another approach that provides a solution for the estimation problem is MHE. The MHE can be viewed as a dual formulation of model predictive control (MPC), and is based on Bayesian *maximum a posteriori* (MAP) estimation. Unlike Kalman-based approaches, the MHE relies on linear programming or nonlinear programming solvers to find an estimate solution at each sampling instant.

4.2.2 Moving horizon estimation algorithm

The Bayesian MAP estimate of the state x given the measurement y is defined as

$$\hat{x} = \arg\max_x p(x|y) \tag{4.4}$$

Eq.4.4 means that the most likely value of x given y should be obtained as the estimate $\hat{x}$. For state space models, we want to find the MAP estimates of the states $\{x_0, \ldots, x_T\}$ given the measurements $\{y_0, \ldots, y_{T-1}\}$, that is,

$$\{\hat{x}_0, \ldots, \hat{x}_T\} = \arg \max_{\{x_0,\ldots,x_T\}} p(x_0, \ldots, x_T | y_0, \ldots, y_{T-1}) \tag{4.5}$$

Using Bayes' rule as well as the nonlinear model of Eq.4.1, one can derive that (see [40] for detail)

$$\begin{aligned}
&\{\hat{x}_0, \ldots, \hat{x}_T\} \\
&= \arg \max_{\{x_0,\ldots,x_T\}} p(x_0, \ldots, x_T | y_0, \ldots, y_{T-1}) \\
&= \arg \min_{\{x_0,\ldots,x_T\}} \Sigma_{k=0}^{T-1} \log p_{v_k}(y_k - h(x_k)) + \log p(x_{k+1}|x_k) + \log p_{x_0}(x_0) \\
&= \arg \min_{\{x_0,\ldots,x_T\}} \Sigma_{k=0}^{T-1} \left(\|y_k - h(x_k)\|^2_{R^{-1}} + \|x_{k+1} - f(x_k, u_k)\|^2_{Q^{-1}} \right) + \|x_0 - \bar{x}_0\|^2_{P_0^{-1}} \\
&= \arg \min_{\{x_0,\ldots,x_T\}} \Sigma_{k=0}^{T-1} \left(\|v_k\|^2_{R^{-1}} + \|w_k\|^2_{Q^{-1}} \right) + \|x_0 - \bar{x}_0\|^2_{P_0^{-1}}
\end{aligned}$$

where w_k and v_k are estimated noise and P_0 and $\bar{x}_0$ are initial guesses of the covariance and mean of the state. Hence, the solution of the Bayesian MAP estimation of states $\{x_0, \ldots, x_T\}$ becomes a minimization problem. This minimization problem can be re-formulated as

$$\arg \min_{x_0, \{w_k\}_{k=0}^{T-1}} \Sigma_{k=0}^{T-1} L_k(w_k, v_k) + \Gamma(x_0) \tag{4.6}$$

where

$$L_k(w_k, v_k) = \|v_k\|^2_{R^{-1}} + \|w_k\|^2_{Q^{-1}}, \quad \Gamma(x_0) = \|x_0 - \bar{x}_0\|^2_{P_0^{-1}}$$

It should be noted that $\Gamma(x_0)$ is an initial penalty function, which summarizes *a priori* knowledge of the initial state ($\Gamma(\bar{x}_0) = 0$ and $\Gamma(x) > 0$ for $x \neq \bar{x}_0$).

All the state estimates can be obtained by repeatedly solving Eq.4.6 as new measurement arrives. This is called *full information estimation*, which has the best theoretical properties in terms of stability and optimality. We should also note that to obtain the estimate $\{x_0, \ldots, x_T\}$, or x_0, $\{w_k\}_{k=0}^{T-1}$, we need to solve an optimization problem. The problem complexity grows at least linearly with the horizon T, which will make the computational load a major problem.

In order to reduce the computational load, we use a moving window of length N and reformulate the estimation problem as

$$\arg \min_{x_0, \{w_k\}_{k=0}^{T-1}} \Sigma_{k=T-N}^{T-1} L_k(w_k, v_k) + \Sigma_{k=0}^{T-N-1} L_k(w_k, v_k) + \Gamma(x_0) \tag{4.7}$$

$$= \arg \min_{z, \{w_k\}_{k=T-N}^{T-1}} \Sigma_{k=T-N}^{T-1} L_k(w_k, v_k) + \mathcal{Z}_{T-N}(z) \tag{4.8}$$

where $\mathcal{Z}_{T-N}(z)$ is called the *arrival cost* at time $T-N$. The arrival cost is a penalty function which summarizes the *a priori* knowledge of the state at time $T-N$. The arrival cost can be interpreted as encapsulating all the knowledge before the moving window. The MHE is illustrated in Fig.4.1.

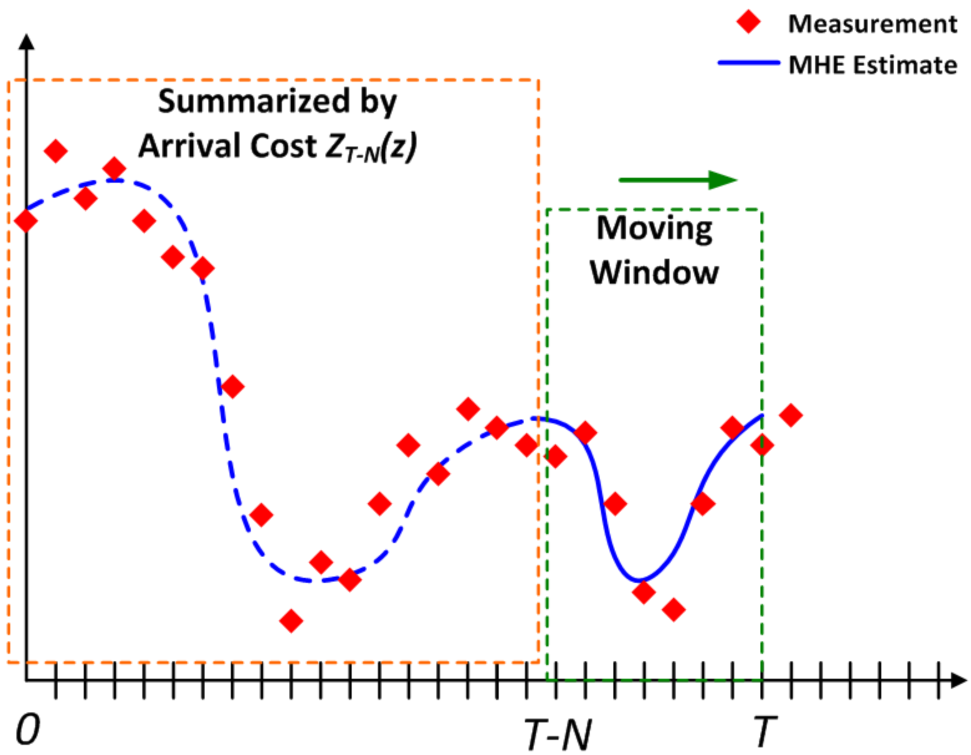


Figure 4.1: Illustration of moving horizon estimation (MHE)

The arrival cost calculation or approximation is important in using MHE and the discussions on arrival cost will be provided later in Section 4.3.

4.2.3 Observability condition of MHE

Observability, which is a system property, is a measure of whether the internal states of a system can be recovered from measurements.

The observability will be determined by the system property if the *full information estimation* in Eq.4.6 is employed to obtain the estimate. However, we normally apply a moving window in the MHE to estimate the unknown states. The prior information outside the moving window will have to be approximated using the arrival cost. If we have a poor arrival cost approximation, then we have to check the observability condition of the MHE in order to recover the states of a system within the window length N.

The observability rank condition for the MHE is that the row rank of $\frac{\partial G}{\partial x}$ must equal n (number of states), where matrix G is defined as [41]

$$G = \begin{bmatrix} h(x_{t-N}) \\ h \circ f(x_{t-N}) \\ \vdots \\ h \circ f \circ \ldots \circ f(x_{t-N}) \end{bmatrix},$$

where $\circ$ is function composition, i.e., $h \circ f = h(f(x))$. It means that we need to ensure that there are sufficient measurements available in the moving window to recover the unknown states. For a linear system, $\frac{\partial G}{\partial x}$ (Eq.4.9) is similar to the observability matrix $\mathcal{O}$ (Eq.4.10). This is shown below:

$$\frac{\partial G}{\partial x} = \frac{\partial \begin{bmatrix} Cx_{t-N} \\ CAx_{t-N} \\ \vdots \\ CA^{N-1}x_{t-N} \end{bmatrix}}{\partial x_{t-N}} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix} \tag{4.9}$$

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{4.10}$$

For a nonlinear system, the observability rank condition only implies local observability around the x. For a comprehensive discussion of observability in MHE, readers are referred to [41].

4.3 Arrival cost approximations in MHE

The major issue with the *full information estimation* shown in Eq.4.6 in real applications lies in the computational load, which makes it hard to be applied *on-line*. In order to reduce the computational load, a horizon is chosen and the arrival cost is employed.

All the information (measurements and estimates) before the moving window is summarized in the arrival cost. Generally, it is difficult to compute the arrival cost $\mathcal{Z}_{T-N}(z)$ in Eq.4.8 exactly. The following two types of arrival cost approximations are often used.

The simplest possible approximation of the arrival cost is to pick $\mathcal{Z}_{T-N}(z)$ as a constant, which means the information before the window is not taken into consideration. By this method, a lower performance will be expected. A simulation will be presented later in Section 4.5.

Another approach, which is the most common approach, is to approximate the arrival cost by

$$\hat{\mathcal{Z}}_{T-N}(z) = \|z - \hat{x}^{mh}_{T-N}\|^2_{P^{-1}_{T-N}} \tag{4.11}$$

where P_k is the variance/covariance of the estimated trajectory $\{\hat{x}^{mh}_k\}$. Eq.4.11 can be seen as applying a penalty for $\{\hat{x}^{mh}_k\}$ (the initial guess in the moving window) when trying to solve an optimization problem. Therefore, a good approximation of the covariance P is important to get a good arrival cost approximation in MHE. From stability considerations, the estimator should not weight the past data too much [42, 41]. This is especially important when comes to the initial guess of P_0.

Linear system

The covariance can be calculated and updated based on the Kalman filter (KF) in the case of a linear system. The linear system is shown as

$$x_{k+1} = Fx_k + Bu_k + w_k \tag{4.12a}$$
$$y_k = Hx_k + v_k \tag{4.12b}$$

In Kalman filter, the prior covariance can be calculated using

$$P_{k+1}^- = FP_kF^T + Q \tag{4.13}$$

with an initial condition given by P_0.

The Kalman gain can be calculated using

$$K = P_k^- H^T (HP_k^- H^T + R)^{-1} \tag{4.14}$$

The posterior covariance can be calculated using

$$P_k^+ = (I - KH)P_k^- \tag{4.15}$$

Then, by substituting Eq.4.14 into Eq.4.15, and also substituting Eq.4.13 into Eq.4.15, we have

$$P_{k+1} = FP_kF^T - FP_kH^T(HP_kH^T + R)^{-1}HP_kF^T + Q \tag{4.16}$$

The covariance propagation for a linear model can be accomplished using Eq.4.16, which is known as the discrete Riccati equation.

Nonlinear system

For nonlinear discrete systems, an approximation of covariance propagation can still be obtained by using Eq.4.16. In the nonlinear application of discrete Riccati equation, F and H are the Jacobians of nonlinear functions $f(\cdot)$ and $h(\cdot)$.

Most systems in practice are continuous processes with discrete measurements. Therefore, the system model and measurement model are given by

$$\dot{x} = f(x, u) + w \tag{4.17a}$$
$$y_k = h(x_k) + v_k \tag{4.17b}$$

In the continuous-discrete system, we can propagate the covariance by using the following two equations:

$$\dot{P}(t) = FP(t) + P(t)F^T + Q \quad \text{where we get } P^- \tag{4.18}$$

$$P_{k+1} = (I - P^-H^T(HP^-H^T + R)^{-1}H)P^- \tag{4.19}$$

where F and H are Jacobians of the nonlinear function $f(\cdot)$ and $h(\cdot)$. The covariance propagation in this method is the same as in the continuous-discrete extended Kalman filter.

Besides using the EKF method to propagate the covariance of a nonlinear system, the unscented Kalman filter (UKF) can also be used to propagate the covariance. A brief procedure for each iteration in the UKF is shown below:

1. Selection of sigma points $\mathcal{X}$.

2. Sigma points propagation through nonlinear model $f(\cdot)$ and $h(\cdot)$.

3. Updating sigma points.

4. Calculation of the posterior covariance of the sigma points.

The UKF provides better covariance approximation than the EKF, and details can be found in Chapter 2.

Moving horizon estimation with larger window sizes will benefit the estimation performance; however, it also becomes the weakness of MHE in real *on-line* applications. It is shown earlier in this section that the arrival cost can be employed to reduce the window size. Once we can get a good approximation for the arrival cost, a larger window may not be necessary. Simulations will be provided later in Section 4.5 to demonstrate the proper selection of window size.

4.4 Inequality constrained parameter estimation with MHE

In practical applications, parameter constraints are commonly encountered. In this work, we consider inequality parameter constraints as

$$d^L \leq c(\theta) \leq d^U \tag{4.20}$$

where d^L and d^U indicate the lower and upper bound of the inequality constraints. Since MHE is an optimization based method, it can take constraints into consideration naturally. In this section, we will first investigate the performance of the constrained MHE. Then, a new constrained parameter estimation method with MHE is proposed.

4.4.1 Problems with directly constrained MHE

In the unconstrained MHE, $z, \{w_k\}_{k=T-N}^{T-1}$ is obtained by solving the following optimization problem:

$$\hat{z}, \{\hat{w}_k\}_{k=T-N}^{T-1} = \arg \min_{z, \{w_k\}_{k=T-N}^{T-1}} \Sigma_{k=T-N}^{T-1} \left(\|v_k\|_{R^{-1}}^2 + \|w_k\|_{Q^{-1}}^2 \right) + \|z - \hat{x}_{T-N}^{mh}\|_{P_{T-N}^{-1}}^2 \tag{4.21}$$

Please note that in Eq.4.21, we have not included constraints into this optimization problem. The arrival cost, or more specifically, the covariance P, can be calculated either by the EKF or the UKF algorithm. Normally, the EKF or the UKF algorithm also does not consider constraints when propagating the covariance.

Then, consider the MHE which takes the constraint into consideration directly. The optimal solution at each sampling instant is obtained using:

$$\tilde{z}, \{\tilde{w}_k\}_{k=T-N}^{T-1} = \arg \min_{z, \{w_k\}_{k=T-N}^{T-1}} \Sigma_{k=T-N}^{T-1} \left(\|v_k\|_{R^{-1}}^2 + \|w_k\|_{Q^{-1}}^2 \right) + \|z - \hat{x}_{T-N}^{mh}\|_{P_{T-N}^{-1}}^2 \tag{4.22}$$

subject to

$$d^L \leq c(\theta) \leq d^U \tag{4.23}$$

As we can see, the only difference between the Eq.4.21 and Eq.4.22 is that the latter is constrained when solving the optimization problem. This also explains why the MHE is able to handle constraints naturally. However, Eq.4.22 may not deliver good estimation performance due to the arrival cost approximation. If the arrival cost is approximated by using the EKF or the UKF, the information of the constraint may not be adequately captured in the arrival cost; hence, the performance is not satisfactory. This problem is illustrated in Fig.4.2.

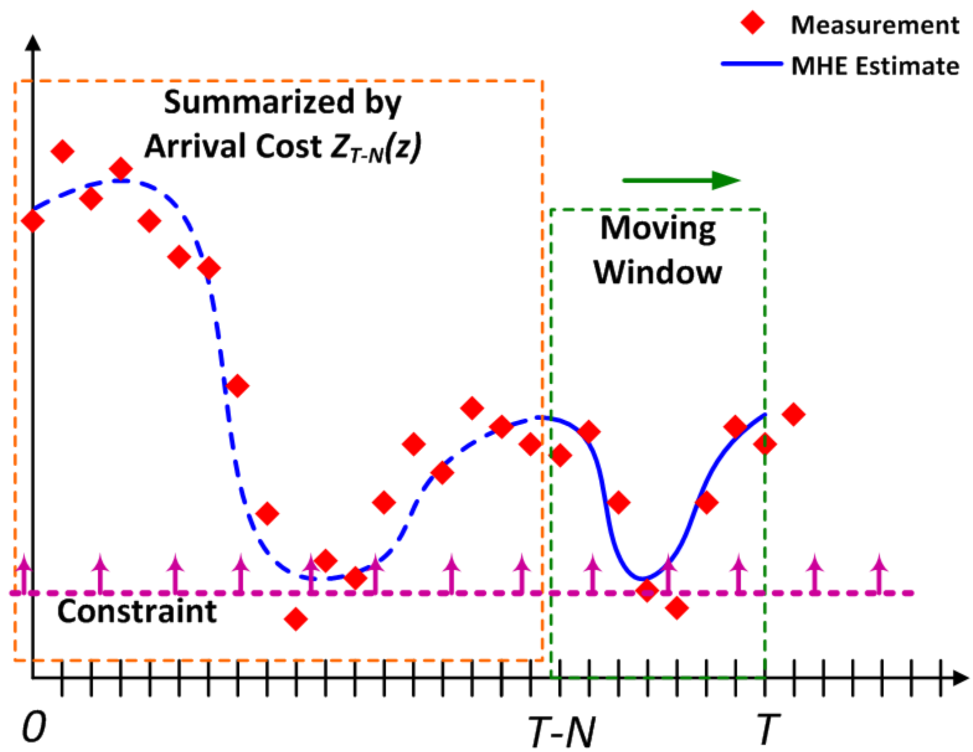


Figure 4.2: Illustration of moving horizon estimation (MHE) with constraints

In Fig.4.2, it is feasible to incorporate the constraints within the moving window. However, the constraint information before the window may not be adequately captured in the arrival cost; hence, the performance is not satisfactory. A general analytical expression for the constrained arrival cost is rarely available, and it has been an open problem [43].

4.4.2 Proposed method

The inequality constraints for parameters can be viewed as constraints on the model. Therefore, the constraints can provide us some knowledge to obtain a better model for estimation. Hence, we propose a two-step method to include the inequality constraints into the estimation with MHE.

In the first step, we use the unconstrained MHE to obtain an unconstrained estimate.

$$\hat{z}, \{\hat{w}_k\}_{k=T-N}^{T-1} = \arg \min_{z,\{w_k\}_{k=T-N}^{T-1}} \Sigma_{k=T-N}^{T-1} \left(\|v_k\|_{R^{-1}}^2 + \|w_k\|_{Q^{-1}}^2 \right) + \|z - \hat{x}_{T-N}^{mh}\|_{P_{T-N}^{-1}}^2 \tag{4.24}$$

Then in the second step, we update the model by using the simple projection method,

$$\tilde{x} = \arg \min_x (x - \hat{x})^T W (x - \hat{x}) \tag{4.25}$$

subject to

$$d^L \leq c(\theta) \leq d^U \tag{4.26}$$

where W is a weighting matrix which can be either I or P^{-1}. If we set $W = I$ we obtain the least squares estimate subject to the constraints. If we set $W = P^{-1}$, it will result in a constrained estimate that is closer to the true value than the unconstrained estimate in each iteration [44].

In the first step of the proposed method, the covariance P is calculated using the EKF or the UKF algorithm. That is to say, the covariance P does not have information of the constraints either. The constraints are only employed in updating the model in Eq.4.25. Even though there is not much information of the constraints in the covariance P, it still can be used to approximate the arrival cost for MHE along with a better process model. Simulation results will be presented later in Section 4.5 showing the efficacy of this method.

4.5 Simulation examples

4.5.1 Example 1: A discrete nonlinear system

The discrete nonlinear model [42] is given by

$$x_{1,k+1} = 0.99x_{1,k} + 0.2x_{2,k} + w_{1,k} \tag{4.27}$$

$$x_{2,k+1} = -0.1x_{1,k} + \frac{0.5x_{2,k}}{1 + x_{2,k}^2} + w_{2,k} \tag{4.28}$$

$$y_k = x_{1,k} - 3x_{2,k} + v_k \tag{4.29}$$

$$\tag{4.30}$$

where the true initial value of state x is [1,2]. w and v are white noises for state and measurement equations respectively.

The initial guess for this model in MHE is given by

$$x_{ini} = \begin{bmatrix} 10 \\ 10 \end{bmatrix} \quad P_{ini} = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$$

For this example, the window size is set as $N = 3$. Therefore, we need to calculate the optimal solution of $\hat{x}_k$, $\hat{w}_k$, $\hat{w}_{k+1}$ and $\hat{w}_{k+2}$ at each sampling instant.

First, we use the MHE to estimate states x_1 and x_2 without considering information before the window. In this case, the objective function in MHE will be

$$\arg \min_{z,\{w_k\}_{k=T-N}^{T-1}} \Sigma_{k=T-N}^{T-1} L_k(w_k, v_k) + \Phi$$

where Φ is a constant value. The estimation performance with constant arrival cost is shown in Fig.4.3 (for state x_1) and Fig.4.4 (for state x_2).

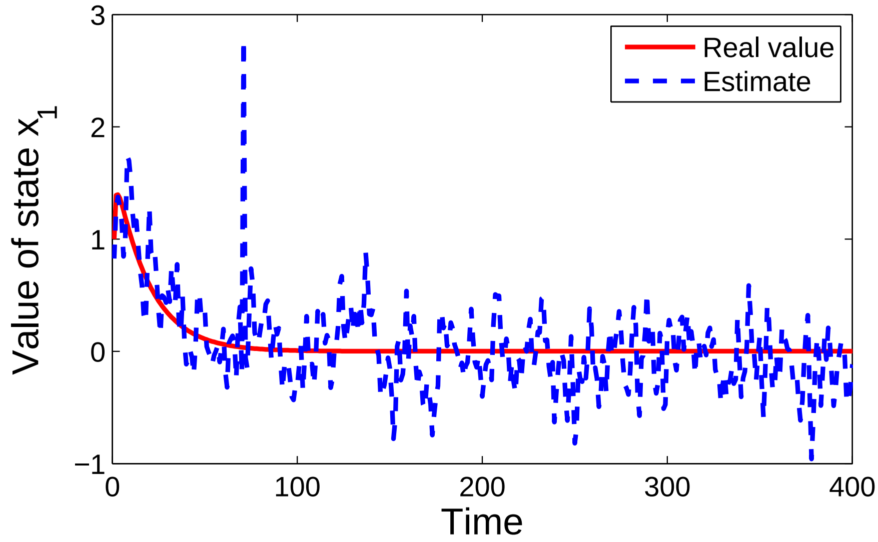


Figure 4.3: Estimation of x_1 with constant arrival cost in MHE for a discrete non-linear model

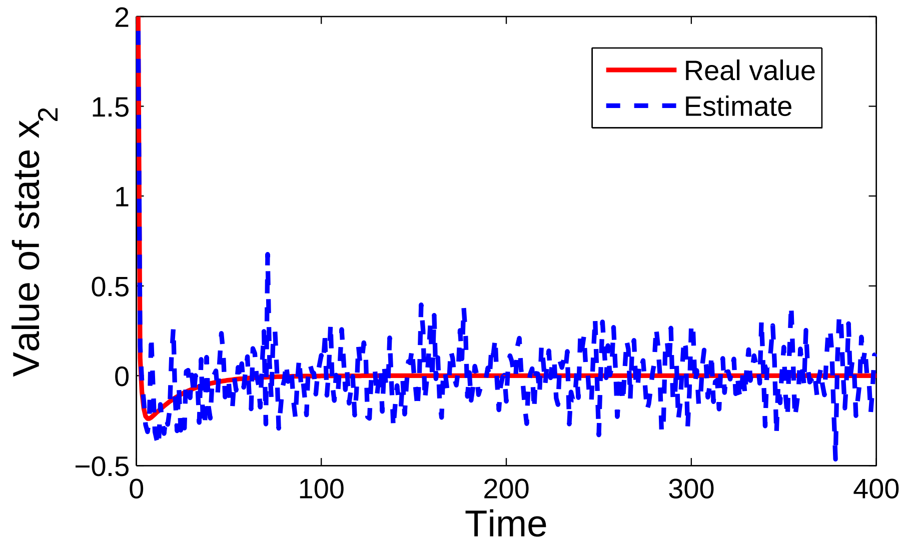


Figure 4.4: Estimation of x_2 with constant arrival cost in MHE for a discrete non-linear model

Then, we consider the arrival cost in the MHE. The objective function of the optimization problem becomes

$$\begin{aligned} &\arg \min_{z,\{w_k\}_{k=T-N}^{T-1}} \Sigma_{k=T-N}^{T-1} L_k(w_k, v_k) + \mathcal{Z}_{T-N}(z) \\ =&\arg \min_{z,\{w_k\}_{k=T-N}^{T-1}} \Sigma_{k=T-N}^{T-1} L_k(w_k, v_k) + \| z - \hat{x}_{T-N}^{mh} \|^2_{P_{T-N}^{-1}} \end{aligned}$$

where P can be calculated using Eq.4.16. The estimation performance is shown in Fig.4.5 (for state x_1) and Fig.4.6 (for state x_2). It can be seen that there is less error in the estimation result compared to the MHE with constant arrival cost. The arrival cost is important and can contribute significantly to the estimation performance in MHE.

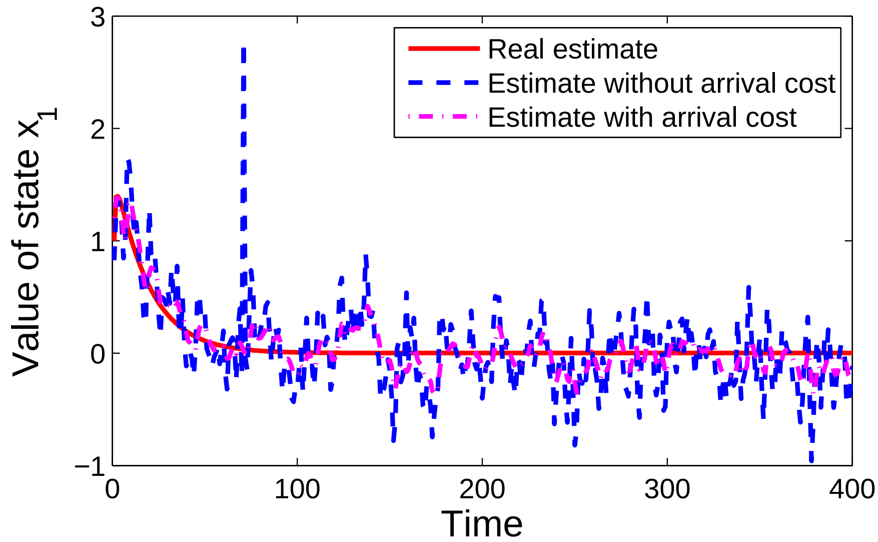


Figure 4.5: Estimation of x_1 with arrival cost approximation in MHE for a discrete nonlinear model

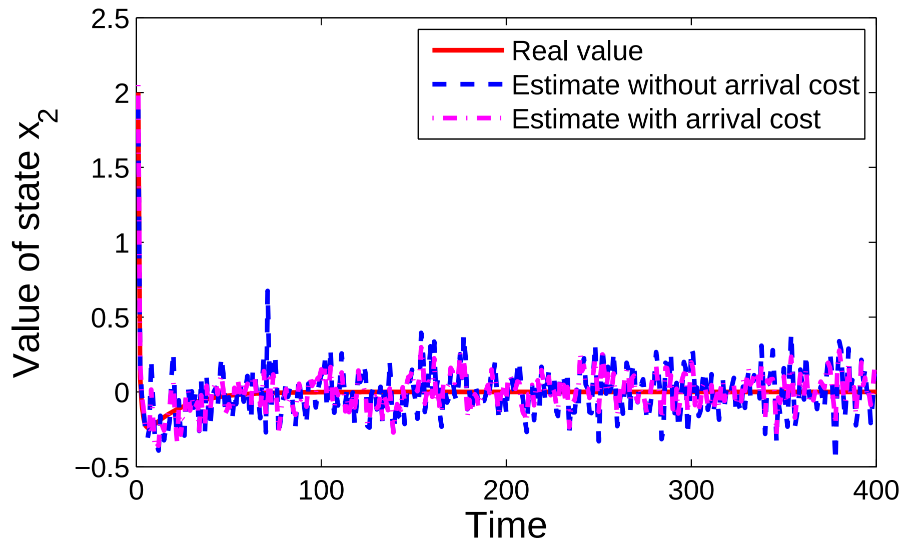


Figure 4.6: Estimation of x_2 with arrival cost approximation in MHE for a discrete nonlinear model

4.5.2 Example 2: CSTR process

CSTR state estimation

The CSTR process involves two state equations and two measurement equations. The CSTR model is shown below:

$$\dot{x_1} = 1 - x_1 - 7.2 \times 10^{10} e^{-\frac{8750}{x_2}} x_1 + w_1 \tag{4.31}$$

$$\dot{x_2} = 350 - x_2 + 150.6276 \times 10^{14} e^{-\frac{8750}{x_2}} \frac{x_1}{1000} + 2092.05 \frac{(u - x_2)}{1000} + w_2 \tag{4.32}$$

$$y_{1,k} = x_{1,k} + v_{1,k} \tag{4.33}$$

$$y_{2,k} = x_{2,k} + v_{2,k} \tag{4.34}$$

with sampling time $T_S = 0.05s$.

Using the observability rank condition, it is shown that the window length $N = 1$ is sufficient for the MHE to be observable for state estimation in this CSTR process. Therefore, the optimal solution will include $\hat{x}_k$, $\hat{w}_k$ at each sampling instant. We use the continuous-discrete EKF method to propagate the covariance in the arrival cost approximation. Simulation results for estimating state x_1 and x_2 are shown in Fig.4.7 and Fig.4.8 respectively. It is shown that the state estimates can converge to their real values with the MHE.

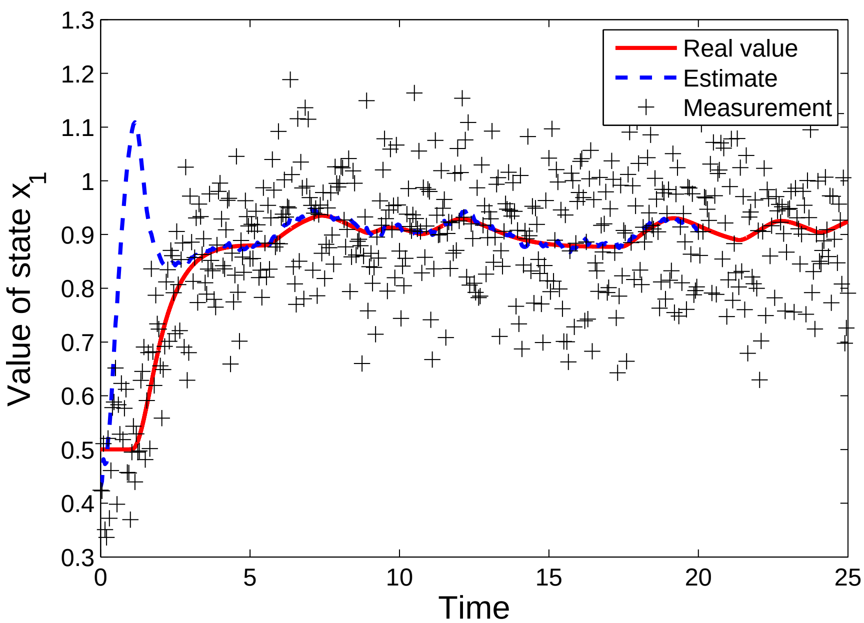


Figure 4.7: State estimation of x_1 in the CSTR model with MHE

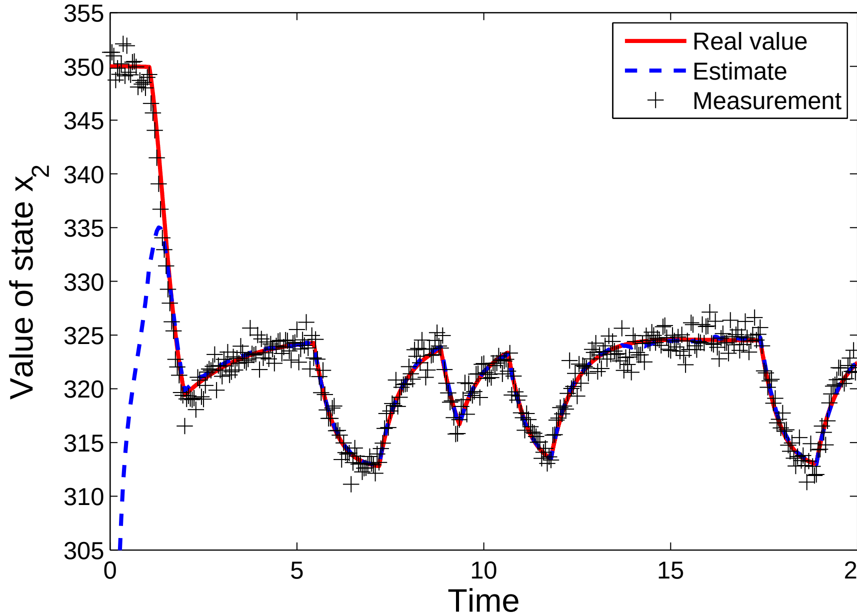


Figure 4.8: State estimation of x_2 in the CSTR model with MHE

CSTR parameter estimation

In this CSTR parameter estimation example, we treat ρ and $\frac{E}{R}$ as augmented states x_3 and x_4 respectively. Hence, the model used in estimation is:

$$\dot{x_1} = 1 - x_1 - 7.2 \times 10^{10} e^{-\frac{x_4}{x_2}} x_1 + w_1 \tag{4.35}$$

$$\dot{x_2} = 350 - x_2 + 150.6276 \times 10^{14} e^{-\frac{x_4}{x_2}} \frac{x_1}{x_3} + 2092.05 \frac{(u - x_2)}{x_3} + w_2 \tag{4.36}$$

$$\dot{x_3} = 0 + w_3 \tag{4.37}$$

$$\dot{x_4} = 0 + w_4 \tag{4.38}$$

$$y_{1,k} = x_{1,k} + v_{1,k} \tag{4.39}$$

$$y_{2,k} = x_{2,k} + v_{2,k} \tag{4.40}$$

with sampling time $T_S = 0.05s$. True values of parameters ρ and $\frac{E}{R}$ are 1000 and 8750 respectively. The observability rank condition shows window length $N = 2$ ensures the MHE to be observable for the states and unknown parameters even with poor arrival cost approximation. Similarly, the optimal solution of $\hat{x}_k$, $\hat{w}_k$ and $\hat{w}_{k+1}$ will be obtained at each sampling instant.

We first use the continuous-discrete EKF method for MHE arrival cost approximation in the CSTR parameter estimation. Then, the UKF method for arrival cost approximation is also applied. Simulation comparisons are shown in Fig.4.9 and Fig.4.10. It can be observed that the UKF is better than the EKF in the arrival cost approximation, because the latter involves large linearization errors.

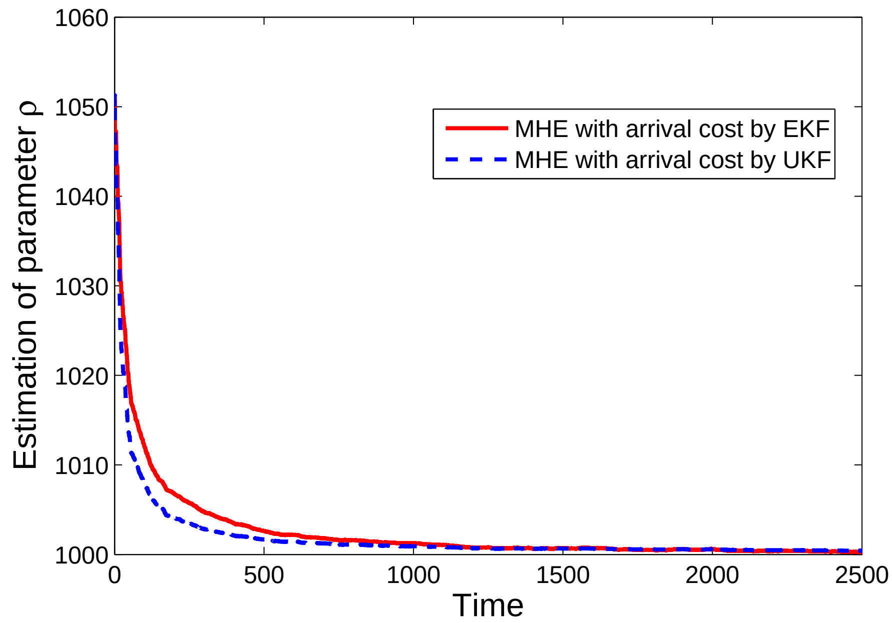


Figure 4.9: Comparison of estimation of parameter ρ in the CSTR model using MHE with different arrival cost approximations

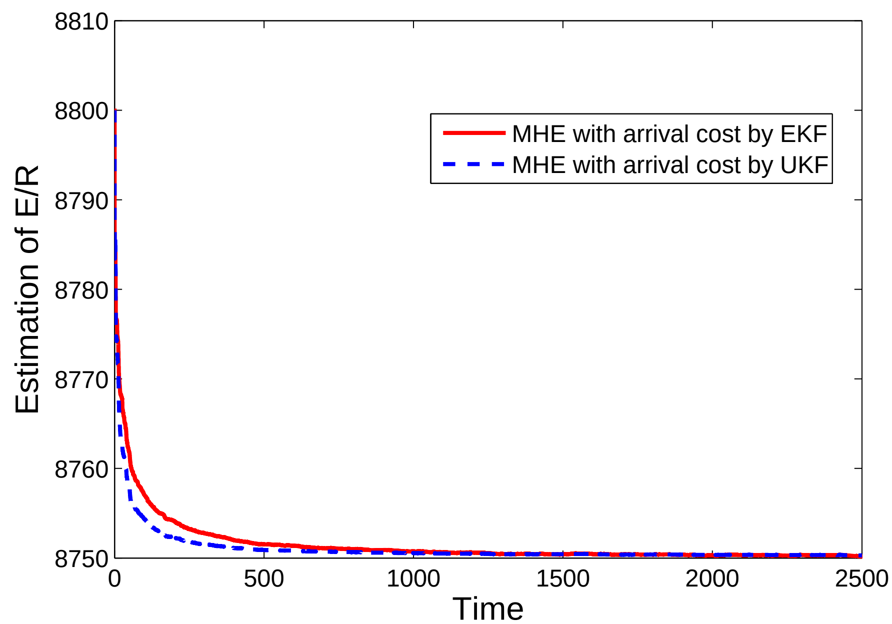


Figure 4.10: Comparison of estimation of parameter $\frac{E}{R}$ in the CSTR model using MHE with different arrival cost approximations

It is said previously that N should be at least 2 in order to ensure the observability of MHE if we have a poor approximation of the arrival cost. Since we have investigate the above approximation methods, we decrease the window length to $N = 1$ and see how it works. Simulation results with a smaller window length of

$N = 1$ are compared with $N = 2$ in Fig.4.11 and Fig.4.12. It is seen that as long as we have good approximation of arrival cost, a smaller window size can retain the accuracy of the state estimation.

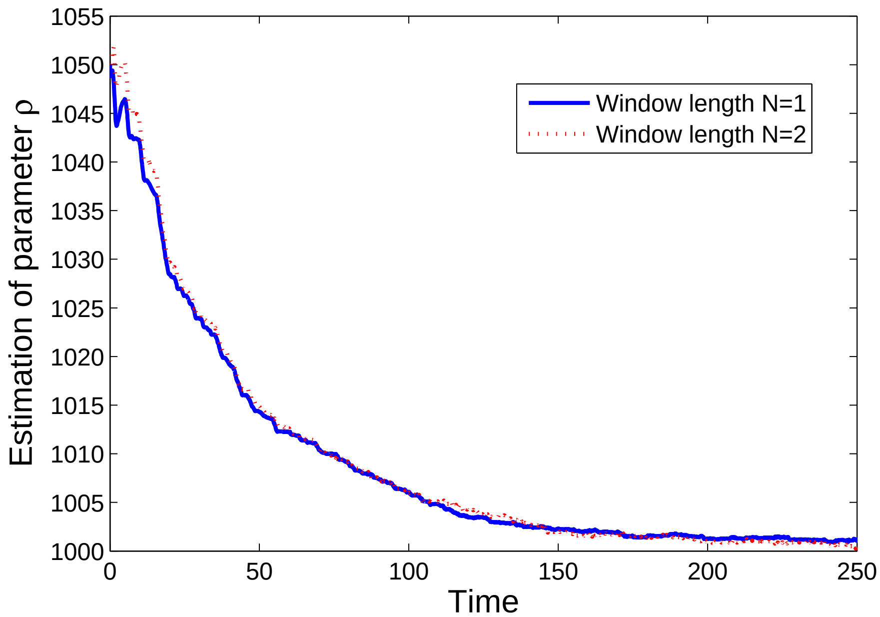


Figure 4.11: Comparison of estimation of parameter ρ in the CSTR model using MHE with different window lengths

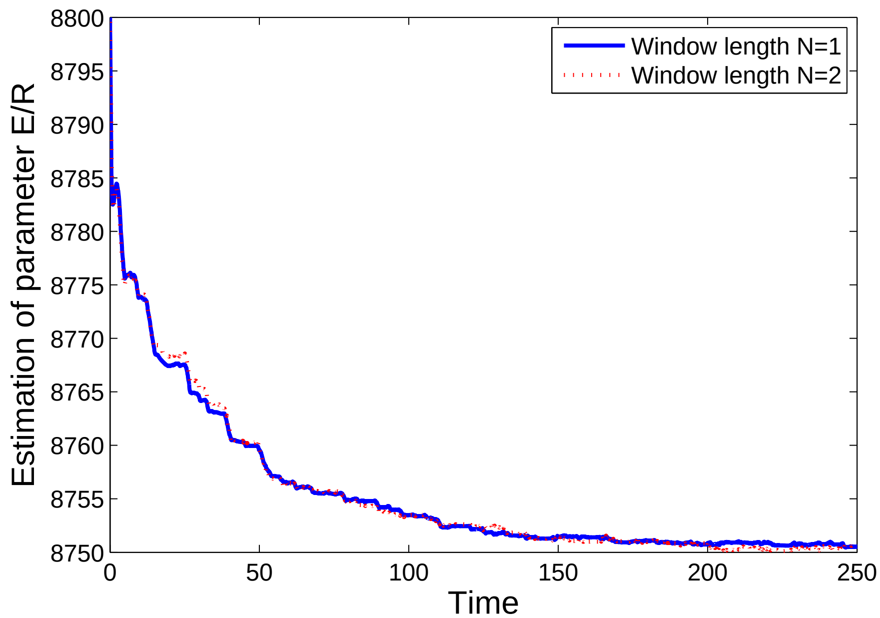


Figure 4.12: Comparison of estimation of parameter $\frac{E}{R}$ in the CSTR model using MHE with different window lengths

CSTR constrained parameter estimation

In this section, we consider the inequality constrained parameter estimation problem of the CSTR model. The methods described in Section 4.4 will be applied to demonstrate their performance. The inequality parameter constraints come from the steady-state measurements of the process and are shown below,

$$\begin{bmatrix} 0.9451 \\ 312.6003 \end{bmatrix} \leq \begin{bmatrix} c_1(\theta) \\ c_2(\theta) \end{bmatrix} \leq \begin{bmatrix} 0.9570 \\ 312.7239 \end{bmatrix} \tag{4.41}$$

Detail of these constraints can be found in the Chapter 2.

First, the MHE in which the constraints are added directly is applied. The UKF is used for the arrival cost approximation. Since the arrival cost approximation is not informative enough to include the information of constraints, it is inaccurate. Hence, there will be large bias in the estimation performance. Fig.4.13 and Fig.4.14 are the simulation results for estimating parameters ρ and $\frac{E}{R}$ with this constrained MHE. The black line indicates unconstrained estimation, while the blue dashed line indicates constrained estimation. It is shown that this method will result in large bias in the estimation.

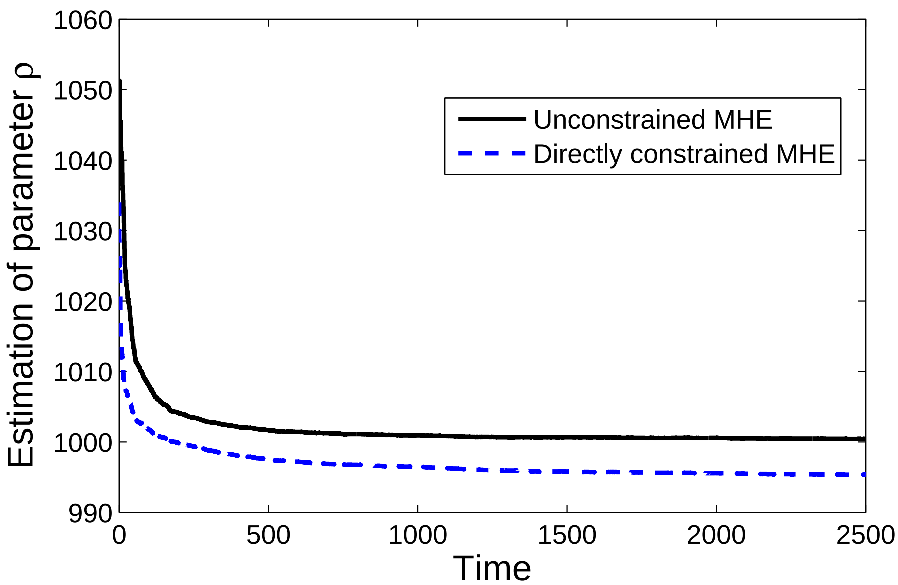


Figure 4.13: Comparison of estimation of parameter ρ in the CSTR model between unconstrained MHE and directly constrained MHE

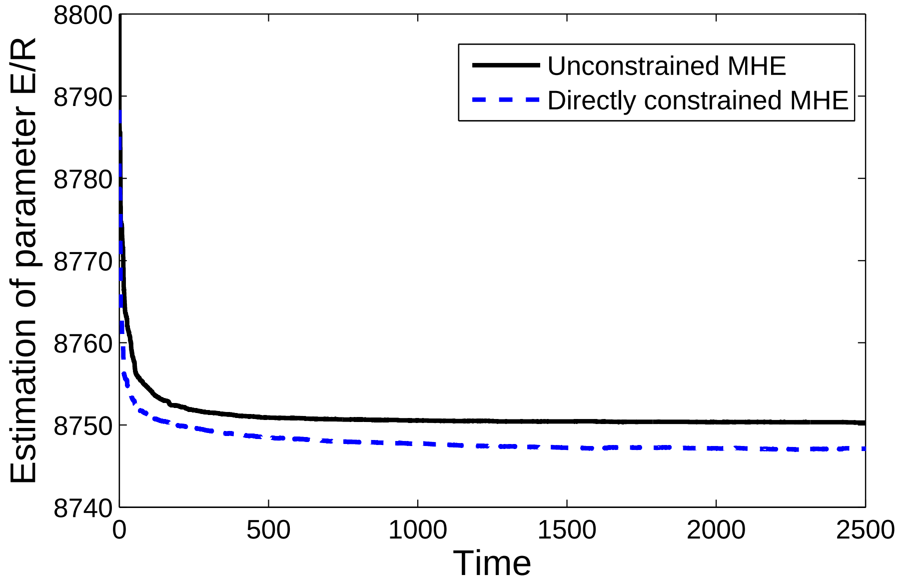


Figure 4.14: Comparison of estimation of parameter $\frac{E}{R}$ in the CSTR model between unconstrained MHE and directly constrained MHE

The proposed method is then applied to the estimation problem. In the proposed method, the model of the process is updated because of the constraints at each sampling instant. Hence, we will have better recovery of the estimates in the MHE. Fig.4.15 and Fig.4.16 are the simulation results for estimating parameters ρ and $\frac{E}{R}$ with the proposed method. It is shown that the proposed method has faster convergence than the unconstrained MHE, and has no bias in the estimation.

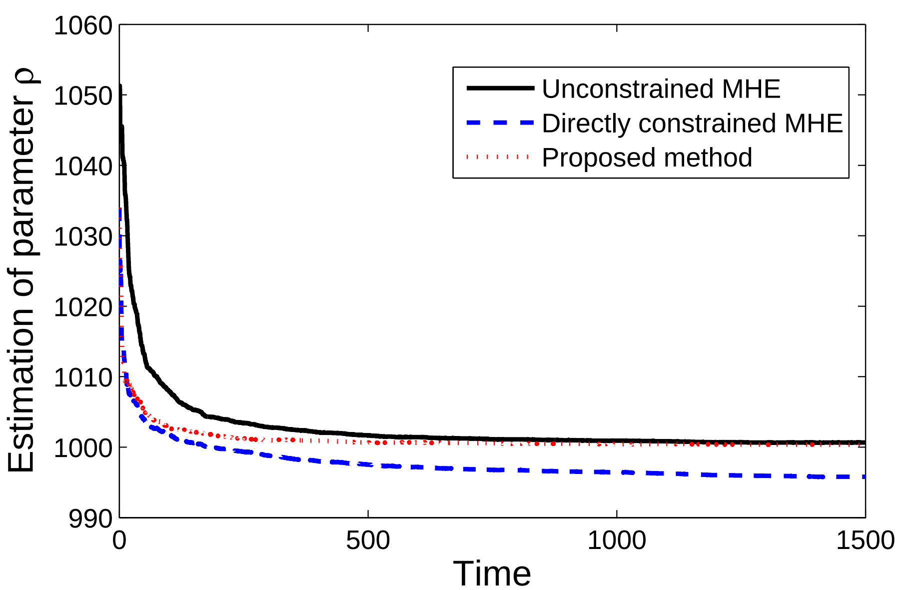


Figure 4.15: Simulation comparison between constrained and unconstrained estimation and the proposed method for estimating parameter ρ in the CSTR model

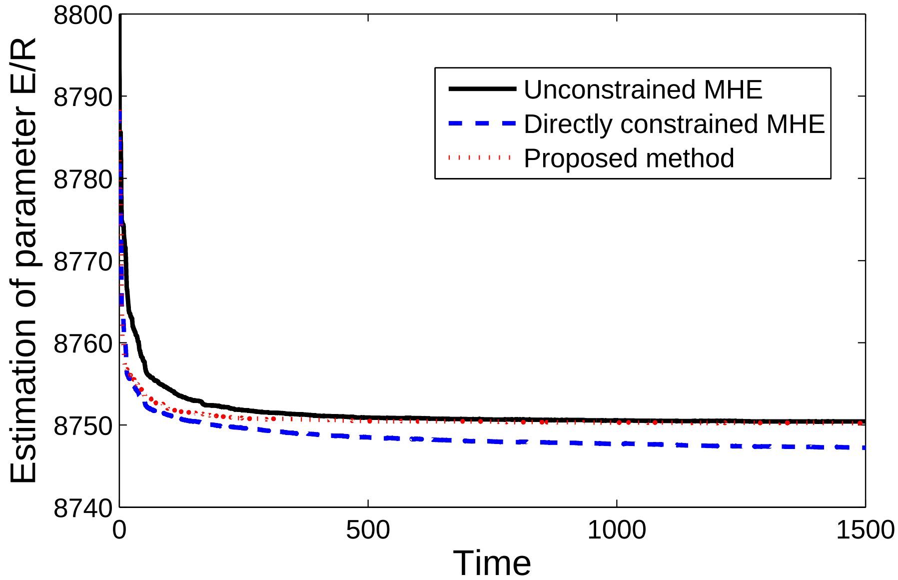


Figure 4.16: Simulation comparison between constrained and unconstrained estimation and proposed method for estimating parameter $\frac{E}{R}$ in the CSTR model

4.6 Conclusions

In this chapter, we have investigated the moving horizon estimation (MHE) method for constrained parameter estimation. Various arrival cost approximations for the unconstrained MHE are presented and evaluated using a discrete nonlinear model as well as a CSTR model. The major problem with the directly constrained MHE is the difficulty of incorporating constraint information into the arrival cost. The proposed method provides an alternative way for constrained parameter estimation with MHE. It can result in faster convergence than the unconstrained MHE and also has better performance than the directly constrained MHE. A continuous-discrete CSTR model with inequality parameter constraints is employed to demonstrate the performance of the proposed method.

Chapter 5

Conclusions and recommendations

5.1 Concluding remarks

The main contributions of this thesis are the development of inequality parameter constraints from steady-state routine operation data, and proposals of constrained estimation methods which result in improved constrained estimation performance. The specific contributions of this thesis can be summarized as follows:

1. Development of a method to construct inequality parameter constraints from steady-state routine operation data.
2. Development of a constrained parameter estimation framework for the UKF.
3. Development of a constraint implementation method under the UKF framework which provides faster convergence than the projection method.
4. Proposal of the appropriate use of projection method in constraining the particles in the EnKF, which retains the same covariance as the unconstrained ensemble.
5. Development of a new constrained parameter estimation method with the EnKF, which results in better performance than the projection method.
6. Development of an alternative method for constrained parameter estimation with the MHE that provides better performance than the directly constrained MHE.
7. Evaluation of the proposed methods using simulated chemical processes.

5.2 Recommendations for future work

Constrained estimation is an interesting and active area of research. In this thesis, we have investigated the inequality constrained parameter estimation problem. We have proposed several possible solutions but a number of problems remain open. The following problems are worthy of further investigations:

1. Mean and covariance inconsistency problem.
   In constrained estimation with that Kalman filter or EKF, there is always a way to constrain the mean value of the estimate. However, the covariance handling in constrained estimation remains a difficult problem and it is an important issue in estimation algorithms. The problem applies to both the UKF and the EnKF, and an appropriate method should be developed to ensure the mean and covariance, which are recovered from the samples, to be consistent even with constraints.

2. MHE arrival cost approximation.
   The poor arrival cost approximation, which does not include the constraint information, could lead to the failure of MHE. It is important to find a good approximation to the arrival cost to incorporate the constraint information.

3. NLP solvers.
   The search for a solution using nonlinear programming solvers is common to encounter in constrained handling, and it is sometimes time consuming. So there is a need to find better NLP solvers that is more efficient in order for on-line implementation.

Bibliography

[1] Y. Zhu and B. Huang. Constrained receding-horizon experiment design and parameter estimation in the presence of poor initial conditions. *AIChE Journal*, 57(10):2808–2820, 2011.

[2] D. Luenberger. Observers for multivariable systems. *IEEE Transactions onAutomatic Control*, 11(2):190–197, 1966.

[3] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[4] J.L. Crassidis and J.L. Junkins. *Optimal Estimation of Dynamic Systems*. Chapman & Hall/CRC, 2004.

[5] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[6] S.J. Julier, J.K. Uhlmann, and H.F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of the American Control Conference*, volume 3, pages 1628 –1632, Jun 1995.

[7] R. Van Der Merwe, A. Doucet, N. De Freitas, and E.A. Wan. The unscented particle filter. *Technical report CUED/F-INFENG/TR 380*, 2000.

[8] E.A. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. In *The IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*, pages 153–158. IEEE, 2000.

[9] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162, 1994.

[10] S. Gillijns, O.B. Mendoza, J. Chandrasekar, B.L.R. De Moor, D.S. Bernstein, and A. Ridley. What is the ensemble Kalman filter and how well does it work? In *American Control Conference*, 2006.

[11] E.L. Haseltine and J.B. Rawlings. Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Industrial & Engineering Chemistry Research*, 44(8):2451–2460, 2005.

[12] W. Wen and H.F. Durrant-Whyte. Model-based multi-sensor data fusion. In *IEEE International Conference on Robotics and Automation*, pages 1720–1726. IEEE, 1992.

[13] P. Vachhani, R. Rengaswamy, V. Gangwal, and S. Narasimhan. Recursive estimation in constrained nonlinear dynamical systems. *AIChE Journal*, 51(3):946–959, 2005.

[14] R. López-Negrete, S.C. Patwardhan, and L.T. Biegler. Constrained particle filter approach to approximate the arrival cost in moving horizon estimation. *Journal of Process Control*, 21(6):909 – 919, 2011.

[15] S. Julier, J. Uhlmann, and H.F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477 –482, mar 2000.

[16] D.G. Robertson, J.H. Lee, and J.B. Rawlings. A moving horizon-based approach for least-squares estimation. *AIChE Journal*, 42(8):2209–2224, 1996.

[17] E.A. Wan and R. van der Merwe. *The Unscented Kalman Filter*, chapter 7, pages 221–280. John Wiley & Sons, Inc., 2002.

[18] A. Romanenko and J.A.A.M. Castro. The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study. *Computers and Chemical Engineering*, 28(3):347 – 355, 2004.

[19] P. Vachhani, R. Rengaswamy, V. Gangwal, and S. Narasimhan. Recursive estimation in constrained nonlinear dynamical systems. *AIChE Journal*, 51(3):946–959, 2005.

[20] P. Vachhani, S. Narasimhan, and R. Rengaswamy. Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation. *Journal of Process Control*, 16(10):1075–1086, 2006.

[21] J. Prakash, S.C. Patwardhan, and S.L. Shah. Constrained nonlinear state estimation using ensemble Kalman filters. *Industrial & Engineering Chemistry Research*, 49(5):2242–2253, 2010.

[22] J. Prakash, S.C. Patwardhan, and S.L. Shah. On the choice of importance distributions for unconstrained and constrained state estimation using particle filter. *Journal of Process Control*, 21(1):3 – 16, 2011.

[23] S.J. Julier. The scaled unscented transformation. In *Proceedings of the 2002 American Control Conference*, volume 6, pages 4555 – 4559 vol.6, 2002.

[24] D. Simon and D.L. Simon. Constrained Kalman filtering via density function truncation for turbofan engine health estimation. *International Journal of Systems Science*, 41(2):159–171, 2010.

[25] S. Sarkka. Unscented Rauch–Tung–Striebel smoother. *IEEE Transactions on Automatic Control*, 53(3):845 –849, april 2008.

[26] M.A. Henson and D.E. Seborg. *Nonlinear Process Control*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1997.

[27] P. Daoutidis, M. Soroush, and C. Kravaris. Feedforward/feedback control of multivariable nonlinear processes. *AIChE Journal*, 36(10):1471–1484, 1990.

[28] P.L. Houtekamer and H.L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.

[29] V.E. Haugen, G. Evensen, and O.M. Johannessen. Indian ocean circulation: an integrated model and remote sensing study. *Journal of Geophysical Research*, 107(C5):11–1, 2002.

[30] P.L. Houtekamer and H.L. Mitchell. Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3269–3289, 2005.

[31] N. Geir, M. Trond, and E. Vefring. Near-well reservoir monitoring through ensemble Kalman filter. In *SPE/DOE Improved Oil Recovery Symposium*, 2002.

[32] H. Li, S.J. Qin, T.T. Tsotsis, and M. Sahimi. Computer simulation of gas generation and transport in landfills: VI Dynamic updating of the model using the ensemble Kalman filter. *Chemical Engineering Science*, 74:69–78, 2012.

[33] J.L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903, 2001.

[34] G. Burgers, P. Jan van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6):1719–1724, 1998.

[35] Y.A. Thomas. Linear quadratic optimal estimation and control with receding horizon. *Electronics Letters*, 11(1):19 –21, 9 1975.

[36] W.H. Kwon, A. M. Bruckstein, and T. Kailath. Stabilizing state-feedback design via the moving horizon method. In *21st IEEE Conference on Decision and Control*, volume 21, pages 234 –239, Dec 1982.

[37] S.S. Jang, B. Joseph, and H. Mukai. Comparison of two approaches to on-line parameter and state estimation of nonlinear systems. *Industrial & Engineering Chemistry Process Design and Development*, 25(3):809–814, 1986.

[38] K.R. Muske, J.B. Rawlings, and J.H. Lee. Receding horizon recursive state estimation. In *American Control Conference, 1993*, pages 900 –904, june 1993.

[39] M. Tyler and M. Morari. Stability of Constrained Moving Horizon Estimation Schemes. Technical report, Automatic Control Laboratory, Swiss Federal Institute of Technology, 1995.

[40] C.V. Rao. *Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems.* PhD thesis, University of Wisconsin-Madison, 2000.

[41] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44(7):1753 – 1765, 2008.

[42] C.V. Rao, J.B. Rawlings, and D.Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, 2003.

[43] C. C. Qu and J. Hahn. Computation of arrival cost for moving horizon estimation via unscented Kalman filtering. *Journal of Process Control*, 19(2):358 – 363, 2009.

[44] D. Simon. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *Control Theory & Applications, IET*, 4(8):1303–1318, 2010.