Coordinated Distributed Moving Horizon State Estimation for Linear Systems

by

Tianrui An

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Process Control

Department of Chemical and Materials Engineering

University of Alberta

Abstract

Industrial chemical plants are typically large-scale systems with a number of processing units or subsystems, which are connected together via material, energy and information flows. The decentralized control frameworks which in general gives suboptimal control performance is normally used for the control of these large-scale chemical processes. With the increasing scales of industrial processes and the interactions between subsystems, it is more challenging to design control systems to achieve the optimal plant operation, as well as to satisfy the increasing requirements on process safety and environmental regulations. In recent years, the distributed framework has been recognized as a promising framework for the control of large-scale systems with interactions. It is shown that the distributed framework has the potential to achieve the centralized framework performance, while maintaining the flexibility of the decentralized control scheme.

This thesis focuses on the development of coordinated distributed state estimation schemes. Specially, we propose coordination algorithms for distributed moving horizon state estimators (MHEs) for discrete-time linear systems. In particular, the class of linear system is composed of several subsystems that interact with each other via their states. Two coordination algorithms are studied: the price-driven coordination algorithm and the prediction-driven coordination algorithm. In the proposed coordinated distributed MHE (CDMHE) schemes, each subsystem is associated with a local MHE. In the design of a local MHE, a coordinating term is incorporated into its cost function which is determined by an upper-layer coordinator. It is shown that both CDMHE schemes are able to achieve the estimation performance of the corresponding centralized design if convergence at each sampling time is ensured.

Acknowledgements

Finally, it is the end and a new beginning. I would like to express my gratitude to those who have supported me and encouraged me during my two-year graduate study.

First of all, I would like to give foremost gratitude to Dr. J. Fraser Forbes and Dr. Jinfeng Liu, who are my dear supervisors. Without their continuous guidance and support, I cannot finish this two-year program. I treasure the meeting time with Fraser, in which everyone is willing to share the recent stories, research progress and even some tourism destinations. Jinfeng is so patient and gives me detailed guidance when I meet difficulties in research. I struggled a lot with this project and almost gave up at one time. That was his support and trust made me finish this thesis. He is also like a friend with whom I can discuss almost everything. I would like to give my appreciation again to my dear supervisors. It was a great pleasure to work under their supervision.

I would like to thank Dr. Padideh Mohseni, Dr. Bardia Hassanzadeh, Dr. Marcos Natalia and Ms. Shuning Li for their work on the coordinated distributed model predictive control algorithms. Especially, I would like to give my sincerely thanks to Bardia Hassanzadeh for his help and advice on my thesis. I would also like to thank my friends here in the CPC group for their support and the joyful time we spent together: Ouyang Wu, Su Liu, Janatan Nahar, Kevin Arulmaran, Wenhan Shen, Jayson McAllister, Qingqing Xu, Peyman Tajik, Chaoqun Li, Yanjun Ma, Ruomu Tan, Mengqi Fang, Xunyuan Yin and so many else. I also wish to give thanks to my friends who make my graduate life so colorful. They are: Saoussen Ouhiba, Lang Liu, Leo Li, Feixia Zhang, Yongzan Liu, Jianfeng Hou, Tianyi Hou, Wenrui Wu, Pengxiang Han, Cong Jin and my dearest housemate Pratya Singhsa.

Last but not least, I would like to express my deepest love and gratitude to my parents, my brother and my family for their unconditional love and support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Chemical plants are typically large-scale systems with a number of processing units or subsystems, which are connected together via material, energy and information flows. The increasing requirements in process safety and environmental regulations, as well as the pursuit of profits and productivity have led to more and more complicated and integrated process designs. Decentralized control frameworks are normally used for the control of these large-scale chemical processes. In the decentralized control frameworks, interactions between subsystems are often ignored or treated in very conservative ways which result in reduced closed-loop performance or even instability. Although centralized control frameworks may give improved (or the optimal) performance, they are not favorable in industries due to computational, organizational and fault tolerance considerations (Bakule, 2008; Christofides *et al.*, 2013).

In the past decade, distributed model predictive control (DMPC) has been recognized as a promising scheme for the control of large-scale systems. Extensive results have been developed for DMPC (e.g.Stewart *et al.*, 2011, Liu *et al.*, 2009, Sun and El-Farra, 2008, Tippett and Bao, 2013, Al-Gherwi *et al.*, 2013, Cheng *et al.*, 2007). A common feature of these DMPC schemes is that the local controllers communicate with each other directly or via a coordinator to exchange information and coordinate their control actions. It is shown that the DMPC has the potential to achieve the centralized control performance, while maintains the flexibility of a decentralized control scheme (Christofides *et al.*, 2013). Reviews of the various existing DMPC results can be found in Christofides *et al.*, 2013, Scattolini, 2009. Pertaining to this work is the coordinated DMPC (CDMPC) framework. The CDMPC framework has a two-layer hierarchical structure. The lower layer local model predictive controllers communicate with an upper-layer coordinator to handle the interactions between the subsystems. In Cheng *et al.*, 2007, a price-driven coordination algorithm was proposed for the optimal control of large-scale linear systems in which a coordinator is used to adjust local controller's behaviors by a price vector which approximates the Lagrange multiplier of the corresponding centralized control problem. In Marcos., 2011, a prediction-driven coordination technique was developed. In this prediction-driven coordination, interactions of subsystems are calculated by the coordinator and sent to local systems. In Mohseni., 2013, the stability and convergence properties of different coordination algorithms were analyzed. While there are extensive results on DMPC, relatively less attention has been given to distributed state estimation within process control.

Regarding distributed state estimation, there are results on decentralized or distributed state estimation which are mainly within two classes: distributed Kalman filtering and distributed moving horizon estimation. The results of distributed Kalman filtering are primarily developed for linear systems in sensor networks (e.g., Khan and Moura, 2008, Mutambara and Durrant-Whyte, 2000, Stanković *et al.*, 2009). One limitation of distributed Kalman filtering schemes is that they cannot take into account constraints. Distributed moving horizon estimation framework (DMHE) has been developed for linear systems (Farina *et al.*, 2010) and then nonlinear systems (Farina *et al.*, 2012, Zhang and Liu, 2013). These DMHE schemes can handle constraints in a systematic way. However, in the design of the above distributed state estimation schemes, no consideration is given to the existing (decentralized) implementation of control/estimation schemes in a process. If a decentralized state estimation scheme has already been implemented in a process, the above distributed schemes essentially require a completely redesign of the existing implementation which means high capital investment.

Inspired by the above considerations, in this thesis, we focus on developing coordinated algorithms for distributed moving horizon state estimators (MHEs) for a class of discrete-time linear systems. If a decentralized MHE scheme is already implemented, the proposed coordination algorithm requires minor modifications to the decentralized scheme and provides significantly improved performance.

1.2 Different Estimation Frameworks

This section provides a brief description of the three state estimation frameworks: the centralized, decentralized and distributed state estimation frameworks.

1.2.1 The Centralized State Estimation Framework

In the centralized state estimation framework, the entire plant is modeled as a whole and the entire plant state is estimated using a centralized observer or estimator, as shown in Figure 1.1. The output measurements of different operating units are all sent to the centralized estimator.

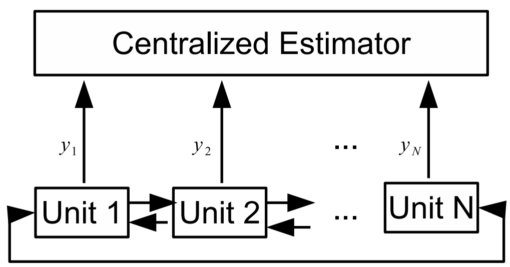

Figure 1.1: A schematic of the centralized state estimation framework

The centralized framework takes into account the interactions between subsystems, and it can give the best possible estimation result. However, with the increasing of the system scale, the computation burden increases significantly. Also, the centralized framework lacks of flexibility and has poor fault tolerance.

1.2.2 The Decentralized State Estimation Framework

In the decentralized state estimation framework, an observer or estimator is designed for each subsystem and is designed based on the corresponding subsystem model. Figure 1.2 shows a schematic of a decentralized design with N subsystems. Since each subsystem has its own estimator, the decentralized framework is easy to implement; however, the interactions between subsystems are typically considered in conservative ways, which lead to suboptimal results or even loss of stability.

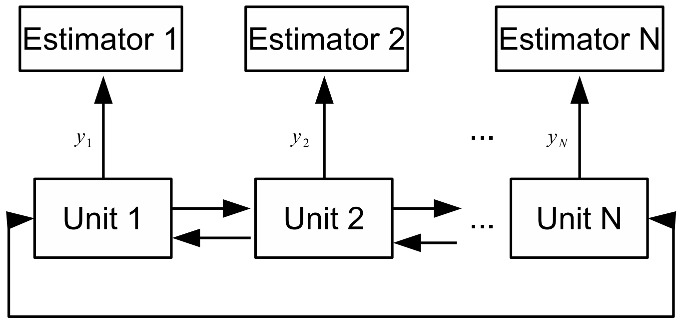

Figure 1.2: A schematic of the decentralized state estimation framework

1.2.3 The Distributed State Estimation Framework

The above concerns for the centralized and decentralized state estimation motivate the development of distributed state estimation schemes. As mentioned in Section 1.1, there are results on distributed state estimation which are mainly within two classes: distributed Kalman filtering and distributed moving horizon estimation. Distributed moving horizon estimation framework (DMHE) has been developed for linear system (Farina *et al.*, 2010) and then nonlinear systems (Farina *et al.*, 2012, Zhang and Liu, 2013). Although existing DMHE schemes vary in the observer formulation and communication structures, the common characteristic they have is that subsystems exchange information with each other. From the information exchange, local estimators get the state estimates of other subsystems and use these information to improve local estimation performance. Figure 1.3 shows a schematic of a distributed state estimation scheme with N estimators.

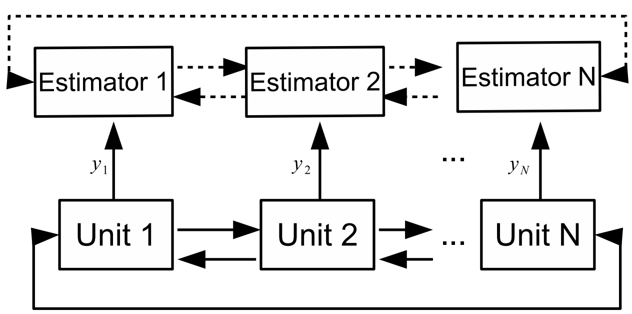

Figure 1.3: A schematic of the distributed state estimation framework

1.3 Thesis Outline and Contributions

Inspired by the coordinated distributed MPC framework in Cheng *et al.*, 2007; Marcos., 2011; Mohseni., 2013, this thesis focuses on the development of coordinated distributed state estimation algorithms. Specifically, the distributed state estimation schemes will be developed in the context of moving horizon estimation. The outline of the thesis and the contributions of each chapter are described below.

Chapter 2 provides a description of the notations, terms as well as the system model and the centralized and decentralized MHE formulations used in this thesis. In Chapter 3, a price-driven coordination algorithm is derived for the distributed moving horizon estimation, where a local MHE estimates all the process states, noises and interactions. It will be shown that the standard price-driven coordination method cannot be used for state estimation purpose since it requires measurements of the full state vector. A improved price-driven CDMHE is proposed to address the issue of the standard version. Firstly, the improved price-driven CDMHE is proposed without considering inequality constraints. Subsequently, a method to handle inequality constraints by dividing the inequality constraints into active constraints and inactive constraints is described. A chemical process example is given to illustrate the effectiveness of the proposed price-driven scheme. In Chapter 4, a prediction-driven CDMHE is presented first without considering inequality constraints. Then a method to handle inequality constraints by using barrier functions is proposed. Two chemical process examples are given to illustrate the applicability and effectiveness of the proposed scheme. In Chapter 5, the performance of the prediction-driven CDMHE is further investigated under different conditions, including triggered communication, communication failure and premature termination. The last chapter, Chapter 6 summarizes the results of this thesis and discusses potential future research directions.

The contributions of this thesis can be summarized as:

- an analysis of the limitation of the standard price-driven coordination algorithm in distributed state estimation;
- a new price-driven coordination algorithm for distributed state estimation;
- a prediction-driven CDMHE with a set of sufficient conditions that ensure the

convergence to the centralized MHE;

- methods to handle inequality constraints in CDMHE schemes;
- examples with extensive simulations illustrating the detailed implementations and demonstrating the performance of the proposed approaches;
- a detailed investigation of the performance of the proposed prediction-driven CDMHE under different conditions.

Chapter 2

Preliminaries

2.1 Terms and Definitions

Some of the key terms used throughout the thesis are explained in this section.

In this thesis, the **plant** or **whole system** indicates the entire system, while **subsystem** means the distributed units that have their own estimators. We use **local** to emphasize the object belongs to subsystems, for example, **local state**, **local estimator**.

MHE stands for moving horizon estimation/estimator. The term **DMHE** denotes distributed MHE, and **CDMHE** refers to coordinated distributed MHE in which DMHEs are coordinated by a **coordinator**. **Price-driven CDMHE** and **prediction-driven CDMHE** refer to the CDMHE network obtained by price-driven coordination method and prediction-driven coordination method, respectively. In the context of **CDMHE**, the **communication** stands for a two-way information exchange between local MHEs and the coordinator. **Iteration** also describes the process of the information exchange and indicates that the information exchange progress is an iterative process.

Centralized MHE means the whole system is estimated by one estimator. **Centralized performance** or **optimal performance** both refer to the optimal performance of the centralized MHE. A **local estimator** refers to the local MHE in a subsystem either in decentralized or distributed estimation network.

2.2 System Description

In this work, the entire plant is assume to be observable and will be described by the following linear, time-invariant, discrete-time system:

$$\boldsymbol{x}(k+1) = A\boldsymbol{x}(k) + \boldsymbol{w}(k); \tag{2.1a}$$

$$\boldsymbol{y}(k) = C\boldsymbol{x}(k) + \boldsymbol{v}(k) \tag{2.1b}$$

where $\boldsymbol{x}(k) \in \mathbb{R}^n$ is the state vector and $\boldsymbol{y}(k) \in \mathbb{R}^q$ is the output vector, $\boldsymbol{w}$ and $\boldsymbol{v}$ represent system and measurement noise terms which have zero means. It is considered that the entire system can be divided into N interconnected subsystems, each with n_i states and q_i outputs, for $i = 1, \ \ldots, \ N$. This implies that $\boldsymbol{x}(k) = \left[\boldsymbol{x}_1(k)^T, \ \boldsymbol{x}_2(k)^T, \ \cdots, \ \boldsymbol{x}_N(k)^T\right]^T$, $\boldsymbol{y}(k) = \left[\boldsymbol{y}_1(k)^T, \ \boldsymbol{y}_2(k)^T, \ \cdots, \ \boldsymbol{y}_N(k)^T\right]^T$, where $\boldsymbol{x}_i \in \mathbb{R}^{n_i}$ and $\boldsymbol{y}_i \in \mathbb{R}^{q_i}$ are the state and output vectors of the i^{th} subsystem, respectively. The system matrix A and output matrix C can be partitioned in the following block-wise fashion:

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{bmatrix}, \tag{2.2}$$

$$C = \begin{bmatrix} C_{11} & & & \\ & C_{22} & & \\ & & \ddots & \\ & & & C_{NN} \end{bmatrix} \tag{2.3}$$

where $A_{ij} \in \mathbb{R}^{n_i \times n_j}$, $C_{ii} \in \mathbb{R}^{n_i \times q_i}$, $i, \ j = 1, \ \ldots, \ N$. The i^{th} subsystem can be represented by the state space model:

$$\boldsymbol{x}_i(k+1) = A_{ii}\boldsymbol{x}_i(k) + \boldsymbol{w}_i(k) + \sum_{i \neq j} A_{ij}\boldsymbol{x}_j(k) \tag{2.4a}$$

$$\boldsymbol{y}_i(k) = C_{ii}\boldsymbol{x}_i(k) + \boldsymbol{v}_i(k) \tag{2.4b}$$

where the part, $A_{ii}\boldsymbol{x}_i(k) + \boldsymbol{w}_i(k)$ in equation (2.4a) denotes the local dynamics of subsystem i. $A_{ij}\boldsymbol{x}_j(k)$, $j \neq i$ represents the coupling between subsystem i and subsystem j.

2.3 The Centralized MHE Formulation

For the system model defined in equation (2.1), we consider a centralized MHE formulated as the following quadratic programming(QP) form:

$$\begin{aligned}\min_{\tilde{X},\tilde{W}} \mathcal{J} &= \frac{1}{2}\sum_{j=k-Hp+1}^{k} \hat{\boldsymbol{v}}(j)' R^{-1}\hat{\boldsymbol{v}}(j) + \frac{1}{2}\sum_{p=k-Hp+1}^{k-1} \hat{\boldsymbol{w}}(p)' Q^{-1}\hat{\boldsymbol{w}}(p); \\ &= \frac{1}{2}\left(\sum_{j=k-Hp+1}^{k} (\boldsymbol{y}(j) - C\hat{\boldsymbol{x}})^T R^{-1}(\boldsymbol{y}(j) - C\hat{\boldsymbol{x}}) + \sum_{p=k-Hp+1}^{k-1} (\hat{\boldsymbol{w}}(p)' Q^{-1}\hat{\boldsymbol{w}}(p))\right);\end{aligned} \tag{2.5a}$$

$$\text{s.t.}\hat{\boldsymbol{x}}(j+1) = A\hat{\boldsymbol{x}}(j) + \hat{\boldsymbol{w}}(j) \tag{2.5b}$$

$$\hat{\boldsymbol{y}}(j) = C\hat{\boldsymbol{x}}(j) + \hat{\boldsymbol{v}}(j) \tag{2.5c}$$

$$\begin{aligned}&\hat{\boldsymbol{x}}(j) \in \mathbb{X},\ \hat{\boldsymbol{w}}(p) \in \mathbb{W} \\ &j = k - Hp + 1, ..., k;\ p = k - Hp + 1, ..., k - 1\end{aligned} \tag{2.5d}$$

where H_p is the estimation horizon. $\tilde{X} = [\hat{\boldsymbol{x}}(k - H_p + 1)^T, \ \cdots, \ \hat{\boldsymbol{x}}(k)^T]^T$, $\tilde{W} = [\hat{\boldsymbol{w}}(k - H_p + 1)^T, \ \cdots, \ \hat{\boldsymbol{w}}(k - 1)^T]^T$ are the estimated state and estimated process noise respectively, $\mathbb{X}$ and $\mathbb{W}$ are constraints on $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{w}}$ respectively, Q and R are assumed to be block diagonal weighting matrices and symmetric positive definite.

Problem (2.5) can be also written in a compact form as follows:

$$\begin{aligned}\min_{\hat{X}(k),\hat{W}(k)} \mathcal{J} =& \frac{1}{2}\begin{bmatrix}\hat{X}(k)^T & \hat{W}(k)^T\end{bmatrix}\begin{bmatrix}\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C} & \\ & \boldsymbol{Q}^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}(k)\\ \hat{W}(k)\end{bmatrix} + \begin{bmatrix}-Y^T\boldsymbol{R}^{-1}\mathbb{C} & 0\end{bmatrix}\begin{bmatrix}\hat{X}(k)\\ \hat{W}(k)\end{bmatrix} \\ =& \frac{1}{2}Z(k)^T \Upsilon Z(k) + \Phi^T Z(k)\end{aligned} \tag{2.6a}$$

$$\begin{aligned}\text{s.t.} \quad & G^{eq}Z(k) = 0 \\ & G^{ineq}Z(k) \le g^{ineq}\end{aligned} \tag{2.6b}$$

where $Z(k) = [\hat{X}(k)^T, \ \hat{W}(k)^T]^T$, $\hat{X}(k) = [\hat{X}_1(k)^T, \ \hat{X}_2(k)^T, \ \cdots, \hat{X}_N(k)^T]^T$, $\hat{W}(k) = [\hat{W}_1(k)^T, \ \hat{W}_2(k)^T, \ \cdots, \ \hat{W}_N(k)^T]^T$, $Y = [Y_1^T, \ Y_2^T, \ \cdots, \ Y_N^T]^T$ with

$$\hat{X}_i(k) = [\hat{\boldsymbol{x}}_i(k - Hp + 1)^T, \ \cdots, \ \hat{\boldsymbol{x}}_i(k)^T]^T, \tag{2.7}$$

$$\hat{W}_i = [\hat{\boldsymbol{w}}_i(k - Hp + 1)^T, \ \cdots, \ \hat{\boldsymbol{w}}_i(k - 1)^T]^T. \tag{2.8}$$

$$Y_i = [\boldsymbol{y}_i(k - H_p + 1)^T, \ \cdots, \ \boldsymbol{y}_i(k)^T]^T \tag{2.9}$$

Note that the formulations of $\hat{X}(k)$ and $\hat{W}(k)$ are different from $\tilde{X}$ and $\tilde{W}$ in (2.5) due to the organization of the subsystem state and process noise.

In (2.6), the definitions of $\mathbb{C}$ and $\boldsymbol{C}_{ii}$ are as follows:

$$\mathbb{C} = \begin{bmatrix} \boldsymbol{C}_{11} & & & \\ & \boldsymbol{C}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{C}_{NN} \end{bmatrix}, \qquad \boldsymbol{C}_{ii} = \begin{bmatrix} C_{ii} & & & \\ & C_{ii} & & \\ & & \ddots & \\ & & & C_{ii} \end{bmatrix}_{H_p \ blocks} \tag{2.10}$$

$\boldsymbol{R}$ and $\boldsymbol{Q}$ are weighting matrix with the following definitions:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 & & & \\ & \boldsymbol{R}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{R}_N \end{bmatrix}, \qquad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_1 & & & \\ & \boldsymbol{Q}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{Q}_N \end{bmatrix} \tag{2.11}$$

with

$$\boldsymbol{Q}_i = \begin{bmatrix} Q_i & & & \\ & Q_i & & \\ & & \ddots & \\ & & & Q_i \end{bmatrix}_{H_p \ blocks}, \qquad \boldsymbol{R}_i = \begin{bmatrix} R_i & & & \\ & R_i & & \\ & & \ddots & \\ & & & R_i \end{bmatrix}_{(H_p-1) \ blocks} \tag{2.12}$$

In equality equation (2.6b), $G^{eq} = [G_A, \ G_B]$, G^{eq} is a $(H_p - 1)n \times (2H_p - 1)n$ matrix, where n is the number of the states in the whole system. G_A and G_B are in the following form:

$$G_A = \begin{bmatrix} G_{A_{11}} & G_{A_{12}} & \cdots & G_{A_{1N}} \\ G_{A_{21}} & G_{A_{22}} & \cdots & G_{A_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ G_{A_{N1}} & G_{A_{N2}} & \cdots & G_{A_{NN}} \end{bmatrix}, \quad G_B = \begin{bmatrix} G_{B_{11}} & G_{B_{12}} & \cdots & G_{B_{1N}} \\ G_{B_{21}} & G_{B_{22}} & \cdots & G_{B_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ G_{B_{N1}} & G_{B_{N2}} & \cdots & G_{B_{NN}} \end{bmatrix} \tag{2.13}$$

with $G_{A_{ii}}$ and $G_{A_{ij}}, i \neq j$ in following forms:

$$G_{A_{ii}} = \begin{bmatrix} -A_{ii} & I_{n_i} & & & \\ & -A_{ii} & I_{n_i} & & \\ & & \ddots & \ddots & \\ & & & -A_{ii} & I_{n_i} \end{bmatrix}_{(H_p-1)n_i \times H_p n_i} \tag{2.14}$$

$$G_{A_{ij}} = \begin{bmatrix} -A_{ij} & & & & 0 \\ & -A_{ij} & & & 0 \\ & & \ddots & & \vdots \\ & & & -A_{ij} & 0 \end{bmatrix}_{(H_p-1)n_i \times H_p n_i} \tag{2.15}$$

where $G_{B_{ii}}$ and $G_{B_{ij}}, i \neq j$ are in following forms:

$$G_{B_{ii}} = \begin{bmatrix} -I_{n_i} & & & \\ & -I_{n_i} & & \\ & & \ddots & \\ & & & -I_{n_i} \end{bmatrix}_{(H_p-1)n_i \times (H_p-1)n_i} \tag{2.16}$$

$$G_{B_{ij}} = 0_{(H_p-1)n_i \times (H_p-1)n_i} \tag{2.17}$$

where G^{ineq} and g^{ineq} in (2.6b) are defined as the following forms:

$$G^{ineq} = \begin{bmatrix} I_{H_pn} & 0 \\ -I_{H_pn} & 0 \\ 0 & I_{(H_p-1)n} \\ 0 & -I_{(H_p-1)n} \end{bmatrix}, \; g^{ineq} = \begin{bmatrix} X_{max} \\ -X_{min} \\ W_{max} \\ -W_{min} \end{bmatrix} \tag{2.18}$$

with $X_{max} = \left[X_{1_{max}}^T, \; \ldots, \; X_{N_{max}}^T\right]^T$, $X_{i_{max}} = [\boldsymbol{x}_{i_{max}}^T, \; \ldots, \; \boldsymbol{x}_{i_{max}}^T]_{H_pn_i}^T$, $X_{min} = [X_{1_{min}}^T, \; \ldots, \; X_{N_{min}}^T]^T$, $X_{i_{min}} = [\boldsymbol{x}_{i_{min}}^T, \; \ldots, \; \boldsymbol{x}_{i_{min}}^T]_{H_pn_i}^T$, $W_{min} = [W_{1_{min}}^T, \; \ldots, \; W_{N_{min}}^T]^T$, $W_{i_{min}} = [\boldsymbol{w}_{i_{min}}^T, \; \ldots, \; \boldsymbol{w}_{i_{min}}^T]_{(H_p-1)n_i}^T$, where n is the total number of states, n_i is the state number of subsystem i, q_i is the output measurement number of subsystem i, I means identity matrix.

2.4 The Decentralized MHE Formulation

For the decentralized MHE, the interactions are not taken into account. For subsystem i, we consider a local MHE is formulated as the following optimization problem:

$$\begin{aligned} \min_{\hat{x}_i, \hat{w}_i} \mathcal{J}_i &= \frac{1}{2} \sum_{j=k-Hp+1}^{k} \hat{\boldsymbol{v}}_i(j)' R_i^{-1} \hat{\boldsymbol{v}}_i(j) + \frac{1}{2} \sum_{p=k-Hp+1}^{k-1} \hat{\boldsymbol{w}}_i(p)' Q_i^{-1} \hat{\boldsymbol{w}}_i(p); \\ &= \frac{1}{2} \left(\sum_{j=k-Hp+1}^{k} (\boldsymbol{y}_i(j) - \boldsymbol{C}_{ii}\hat{\boldsymbol{x}}_i(j))^T R^{-1} (\boldsymbol{y}_i(j) - \boldsymbol{C}_{ii}\hat{\boldsymbol{x}}_i(j)) + \sum_{p=k-Hp+1}^{k-1} (\hat{\boldsymbol{w}}_i(p)' Q_i^{-1} \hat{\boldsymbol{w}}_i(p)) \right); \end{aligned} \tag{2.19a}$$

$$s.t. \; \hat{\boldsymbol{x}}_i(j+1) = A_{ii}\hat{\boldsymbol{x}}_i(j) + \hat{\boldsymbol{w}}_i(j) \tag{2.19b}$$

$$\hat{\boldsymbol{y}}_i(j) = C_{ii}\hat{\boldsymbol{x}}_i(j) + \hat{\boldsymbol{v}}_i(j) \tag{2.19c}$$

$$\begin{aligned} &\hat{\boldsymbol{x}}_i(j) \in \mathbb{X}_i, \; \hat{\boldsymbol{w}}_i(p) \in \mathbb{W}_i \\ &j = k - Hp + 1, ..., k; \; p = k - Hp + 1, ..., k - 1 \end{aligned} \tag{2.19d}$$

where $\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{w}}_i, \hat{\boldsymbol{v}}_i$ are the estimated state, estimated process noise and estimated measurement noise for subsystem i, respectively, R_i and Q_i are weighting matrices, $\boldsymbol{y}_i$ is the provided measurement of subsystem i.

The problem (2.19) can be rearranged in the compact form as follows:

$$
\begin{aligned}
\min_{\hat{X}_i(k),\hat{W}_i(k)} \mathcal{J}_i =& \frac{1}{2}\begin{bmatrix}\hat{X}_i(k)^T & \hat{W}_i(k)^T\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & \\ & \boldsymbol{Q}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} + \begin{bmatrix}-Y_i^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & 0\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix}\\
=& \frac{1}{2}Z_i(k)^T\Upsilon_i Z_i(k) + \Phi_i^T Z_i(k)
\end{aligned} \tag{2.20a}
$$

$$
\text{s.t.} \quad \begin{aligned} & G_i^{eq} Z_i(k) = 0 \\ & G_i^{ineq} Z_i(k) \leq g_i^{ineq} \end{aligned} \tag{2.20b}
$$

where $G_i^{eq} = [G_{A_{ii}},\ G_{B_{ii}}]$, $G_{A_{ii}}$and $G_{B_{ii}}$ are defined in (2.14) and (2.16), respectively. G_i^{ineq} and g_i^{ineq} are defined as follows:

$$
G_i^{ineq} = \begin{bmatrix} I_{H_p n_i} & 0 \\ -I_{H_p n_i} & 0 \\ 0 & I_{(H_p-1)n_i} \\ 0 & I_{(H_p-1)n_i} \end{bmatrix}, \qquad g_i^{ineq} = \begin{bmatrix} X_{i_{max}} \\ -X_{i_{min}} \\ W_{i_{max}} \\ -W_{i_{min}} \end{bmatrix} \tag{2.21}
$$

where $X_{i_{max}},\ X_{i_{min}},\ W_{i_{max}},\ W_{i_{min}}$ are defined the same as in (2.18), n_i is the state number of subsystem i, I means identity matrix.

For the decentralized MHE formulation, the interactions that exists between subsystems are not considered. Therefore the estimated results that obtained from decentralized MHE formulation are always suboptimal.

2.5 Existence of Solution to the MHE

In Section 2.3 and Section 2.4, the arrival cost is not considered in the MHE formulations. In this section, we will focus on the centralized MHE formulation and show that if the estimation horizon satisfies $H_p \geq n$, then the MHE optimization problem is well-posed in the sense that it has a unique (optimal) solution each sampling time.

In order to simplify the analysis, let us focus on a specific case first. Let us choose the estimation horizon $H_p = 3$, the number of states $n = 2$, and number of measurements $p = 1$. Within the estimation horizon, the estimated states of (2.1) satisfy the following recursive equations:

$$
\begin{aligned}
\hat{x}(0) &= \begin{bmatrix} I & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x}(0) \\ \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
\hat{x}(1) &= \begin{bmatrix} A & I & 0 \end{bmatrix} \begin{bmatrix} \hat{x}(0) \\ \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
\hat{x}(2) &= \begin{bmatrix} A^2 & A & I \end{bmatrix} \begin{bmatrix} \hat{x}(0) \\ \hat{w}(0) \\ \hat{w}(1) \end{bmatrix}
\end{aligned} \tag{2.22}
$$

Based on these equations and the objective function in (2.5), it can be obtained that:

$$
\begin{aligned}
\min_{\hat{x}(0),\hat{w}(0),\hat{w}(1)} \mathcal{J} &= \frac{1}{2}\{\sum_{j=0}^{2} \|y(j) - C\hat{x}(j)\|^2_{R^{-1}} + \sum_{q=0}^{1} \|\hat{w}(q)\|^2_{Q^{-1}}\}; \\
&= \frac{1}{2}\{\hat{x}(0)^T C^T R^{-1} C\hat{x}(0) + \hat{x}(1)^T C^T R^{-1} C\hat{x}(1) + \hat{x}(2)^T C^T R^{-1} C\hat{x}(2)\} \\
&+ \frac{1}{2}\{\hat{w}(0)^T Q^{-1}\hat{w}(0) + \hat{w}(1)^T Q^{-1}\hat{w}(1)\} \\
&+ \frac{1}{2}\{y(0)^T R^{-1} y(0) + y(1)^T R^{-1} y(1) + y(1)^T R^{-1} y(1)\} \\
&- \{y(0)^T R^{-1} C\hat{x}(0) + y(1)^T R^{-1} C\hat{x}(1) + y(2)^T R^{-1} C\hat{x}(2)\}
\end{aligned} \tag{2.23a}
$$

$$
s.t.\ \hat{x}(0) \in \mathbb{X},\ \ \hat{w}(0), \hat{w}(1) \in \mathbb{W} \tag{2.23b}
$$

where $\mathbb{X}$ and $\mathbb{W}$ are constraints on $\hat{x}$ and $\hat{w}$, respectively.

Substituting equation (2.22) into equation (2.23) and denoting $r = C^T R^{-1} C$, $l = R^{-1}C$, $z = [\hat{x}(0)^T, \hat{w}(0)^T, \hat{w}(1)^T]^T$ and taking into account that $\frac{1}{2}\{y(0)^T R^{-1} y(0) + y(1)^T R^{-1} y(1) + y(1)^T R^{-1} y(1)\}$ is a constant, the above optimization problem can be rewritten as follows:

$$
\begin{aligned}
\min_{z} \mathcal{J} &= \frac{1}{2}\{z^T\{\begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix} r \begin{bmatrix} I & 0 & 0 \end{bmatrix} + \begin{bmatrix} A^T \\ I \\ 0 \end{bmatrix} r \begin{bmatrix} A & I & 0 \end{bmatrix} + \begin{bmatrix} (A^2)^T \\ A^T \\ I \end{bmatrix} r \begin{bmatrix} A^2 & A & I \end{bmatrix}\}z\} \\
&- \{y(0)^T l' \begin{bmatrix} I & 0 & 0 \end{bmatrix} + y(1)^T l' \begin{bmatrix} A & I & 0 \end{bmatrix} + y(2)^T r' \begin{bmatrix} A^2 & A & I \end{bmatrix}\}z \\
&+ \frac{1}{2} z^T \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q^{-1} & 0 \\ 0 & 0 & Q^{-1} \end{bmatrix} z \\
&= \frac{1}{2} z^T \Upsilon z - \Phi z
\end{aligned} \tag{2.24a}
$$

$$s.t. \quad G_{exp}^{ineq} z \leq g_{exp}^{ineq} \tag{2.24b}$$

where in equation (2.24), Υ, Φ are as follows:

$$\Upsilon = \begin{bmatrix} r + A^T r A + (A^2)^T r A^2 & A^T r + (A^2)^T r A & (A^2)^T r \\ rA + A^T r A^2 & r + A^T r A + Q^{-1} & A^T r \\ rA^2 & rA & r + Q^{-1} \end{bmatrix}$$
$$\Phi = \{y(0)^T l \begin{bmatrix} I & 0 & 0 \end{bmatrix} + y(1)^T l \begin{bmatrix} A & I & 0 \end{bmatrix} + y(2)^T l \begin{bmatrix} A^2 & A & I \end{bmatrix}\} \tag{2.25}$$

where G_{exp}^{ineq} and g_{exp}^{ineq} have the following definitions:

$$G_{exp}^{ineq} = \begin{bmatrix} I_n & 0 & 0 \\ -I_n & 0 & 0 \\ 0 & I_q & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & I_q \\ 0 & 0 & -I_q \end{bmatrix}, \quad g_{exp}^{ineq} = \begin{bmatrix} x_{max} \\ -x_{min} \\ w_{max} \\ -w_{min} \\ w_{max} \\ -w_{min} \end{bmatrix} \tag{2.26}$$

where n is the number of states, q is the number of measurements, x_{max}, x_{min}, w_{max}, w_{min} are upper and lower bounds on $\hat{x}$ and $\hat{w}$, respectively.

Since $\hat{w}(0)^T[rA + A^T r A^2]\hat{x}(0)$ is a scalar, the transpose operation does not change the value. Therefore, $\hat{w}(0)^T[rA + A^T r A^2]\hat{x}(0) = \hat{x}(0)^T[A^T r + (A^2)^T r A]\hat{w}(0)$. Therefore the equation (2.24) can be rewritten as the following form:

$$\min_z \mathcal{J} = \frac{1}{2} z^T \Gamma z - \Phi z$$
$$s.t. \; G_{exp}^{ineq} z \leq g_{exp}^{ineq} \tag{2.27}$$

where

$$\Gamma = \begin{bmatrix} r + A^T r A + (A^2)^T r A^2 & 0 & 0 \\ 2(rA + A^T r A^2) & r + A^T r A + Q^{-1} & 0 \\ 2rA^2 & 2rA & r + Q^{-1} \end{bmatrix} \tag{2.28}$$

The Karush-Kuhn-Tucker (KKT) conditions for a regular point z^* to be an minimum in the quadratic problem (2.27) are:

$$\Gamma z^* - \Phi + G_{exp}^{ineq\,T} \mu^* = 0 \tag{2.29a}$$
$$G_{exp}^{ineq} z^* - g_{exp}^{ineq} \leq 0 \tag{2.29b}$$
$$\mu^{*T}(G_{exp}^{ineq} z^* - g_{exp}^{ineq}) = 0 \tag{2.29c}$$
$$\mu^* \geq 0 \tag{2.29d}$$

where z^* is the optimal solution, μ^* is the Lagrange multiplier.

In order to get unique solution of equation (2.29), the invertibility of matrix Γ is needed; that is $det(\Gamma) \neq 0$.

Lemma 2.5.1. *(Silvester, 2000) Suppose A, B, C and D are matrices of dimension $n \times n$, $n \times m$, $m \times n$, and $m \times m$, resepctively. Then*

$$det\{\begin{bmatrix} A & 0 \\ C & D \end{bmatrix}\} = det(A) \times det(D) = det\{\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}\} \tag{2.30}$$

Based on the above lemma, the determinant of Γ is:

$$det(\Gamma) = det(r + A^T rA + (A^2)^T rA^2)det(r + A^T rA + Q^{-1})det(r + Q^{-1}) \tag{2.31}$$

Because Q^{-1} is positive definite, r and $A^T rA$ are both semi-positive definite, the determinant of the last two matrices are not zero. We just need to show that $det(r + A^T rA + (A^2)^T rA^2) \neq 0$. Let us use P to represent this matrix.

$$\begin{aligned} P &= r + A^T rA + (A^2)^T rA^2 \\ &= \begin{bmatrix} C^T & A^T C^T \end{bmatrix} \begin{bmatrix} R^{-1} & \\ & R^{-1} \end{bmatrix} \begin{bmatrix} C \\ CA \end{bmatrix} + (A^2)^T C^T R^{-1} CA^2 \\ &= o^T \begin{bmatrix} R^{-1} & \\ & R^{-1} \end{bmatrix} o + (A^2)^T C^T R^{-1} CA^2 \end{aligned} \tag{2.32}$$

where $o = \begin{bmatrix} C \\ CA \end{bmatrix}$, is the observability matrix of the system.

Lemma 2.5.2. *(Mirsky, 2012) For a real matrix A, $rank(A^T A) = rank(AA^T) = rank(A) = rank(A^T)$.*

Matrix $\begin{bmatrix} R^{-1} & \\ & R^{-1} \end{bmatrix}$ is the aggregated form of weighting matrix R^{-1}. Since R^{-1} is a full rank and positive symmetric matrix, and usually be picked as a constant multiple identity matrix, we can treat $\begin{bmatrix} R^{-1} & \\ & R^{-1} \end{bmatrix}$ as identity matrix. Then we only need to consider the rank of $o^T o$. Since system is observable, rank(o) $= 2$, thus rank($o^T o$) $= 2$ which is full rank. Therefore, the determinant of P is not zero, thus P is invertible. This implies that the MHE optimization problem has a unique solution.

Now, let us consider the general case with system matrices A with dimensions $n \times n$, C with dimension $m \times n$, and moving horizon H_p. Following similar procedures, it can be obtained that the corresponding P matrix can be obtained as follows:

$$P = r + A^T r A + (A^2)^T r A^2 + \cdots + (A^{H_p-1})^T r A^{H_p-1} \tag{2.33}$$

When estimation horizon $H_p \geq n$, P can be rewritten as:

$$\begin{aligned}
P = & \begin{bmatrix} C^T & A^T C^T & \cdots & (A^{(n-1)})^T C^T \end{bmatrix} \begin{bmatrix} R^{-1} & & & \\ & R^{-1} & & \\ & & \ddots & \\ & & & R^{-1} \end{bmatrix} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{(n-1)} \end{bmatrix} \\
& + \sum_{k=n}^{Hp-1} (A^k)^T C^T R^{-1} C(A^k) \qquad (2.34) \\
= & \vartheta^T \begin{bmatrix} R^{-1} & & & \\ & R^{-1} & & \\ & & \ddots & \\ & & & R^{-1} \end{bmatrix}_{nm*nm} \vartheta + \sum_{k=n}^{Hp-1} (A^k)^T C^T R^{-1} C(A^k)
\end{aligned}$$

where $\vartheta = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{(n-1)} \end{bmatrix}$, is the observability matrix of the system, $\text{rank}(\vartheta) = n$, which implies that determinant of P is not 0. Thus determinant of Γ is not zero and Γ is an invertible matrix. This means that a unique solution exists for the MHE formulation used in Section 2.3 and Section 2.4.

In summary, for an observable system consists of n states, if the window size of the horizon $H_p \geq n$, the MHE formulation used in in Section 2.3 and Section 2.4 has the unique solution.

2.6 Conclusions

In this chapter, some terms and definitions used in this thesis are introduced, the formulations of the centralized MHE and the decentralized MHE are given. It is also shown that such MHE formulation has a unique solution when the estimation horizon is greater than or equal to the number of states.

Chapter 3

Price-driven Coordinated Distributed MHE

In this chapter, a coordination method directly adopted from price-driven CDMPC (Cheng *et al.*, 2007) is first developed for distributed MHE. We then show that this conventional price-driven coordination algorithm cannot be used for state estimation since it requires measurement of the entire system state. Subsequently, an improved price-driven CDMHE is proposed to address this issue. The convergence analysis of the improved price-driven CDMHE under both unconstrained and constrained conditions are given.

3.1 Price-driven Coordination Algorithm

In the proposed price-driven CDMHE, we estimate not only the states $\hat{x}$, process noise $\hat{w}$ but also the interaction $\hat{h}$. The proposed price-driven CDMHE has a two-layer hierarchical structure, in which the local subsystems/estimators are in the lower layer and the coordinator is in the upper layer. We first focus on the unconstrained case. The objective function of the centralized MHE can be rewritten in the summation of N subsystems as follows :

$$\min_{\hat{X}_i(k),\hat{W}_i(k)} \mathcal{J} = \sum_{i=1}^{N} \{ \frac{1}{2} \begin{bmatrix} \hat{X}_i(k)^T & \hat{W}_i(k)^T \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} & \\ & \boldsymbol{Q}_i^{-1} \end{bmatrix} \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \end{bmatrix} + \begin{bmatrix} -Y_i^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} & 0 \end{bmatrix} \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \end{bmatrix} \} \tag{3.1a}$$

$$\text{s.t.} \quad G_{A_{ii}} \hat{X}_i(k) + G_{B_{ii}} \hat{W}_i(k) = \hat{H}_i(k) \tag{3.1b}$$

$$\hat{H}_i(k) = \sum_{j=1, j\neq i}^{N} -G_{A_{ij}}\hat{X}_j(k) \tag{3.1c}$$

The Lagrange function of the problem (3.1) is written in the following form:

$$\begin{aligned}
\mathcal{L} =& \sum_{i=1}^{N}\{\frac{1}{2}\begin{bmatrix}\hat{X}_i(k)^T & \hat{W}_i(k)^T\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & \\ & \boldsymbol{Q}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} + \begin{bmatrix}-Y_i^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & 0\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix}\} \\
&+ \sum_{i=1}^{N}\mu_i^T(G_{A_{ii}}\hat{X}_i(k) + G_{B_{ii}}\hat{W}_i(k) - \hat{H}_i(k)) + \sum_{i=1}^{N}\lambda_i^T(\hat{H}_i(k) + \sum_{j=1, j\neq i}^{N} G_{A_{ij}}\hat{X}_j(k)) \\
=& \sum_{i=1}^{N}\mathcal{F}_i(\hat{X}_i, \hat{W}_i) + \sum_{i=1}^{N}\mu_i^T(G_{A_{ii}}\hat{X}_i + G_{B_{ii}}\hat{W}_i - \hat{H}_i) + \sum_{i=1}^{N}\lambda_i^T(\hat{H}_i + \sum_{j=1, j\neq i}^{N} G_{A_{ij}}\hat{X}_j)
\end{aligned} \tag{3.2}$$

where the time instant index k is omitted in $\hat{X}_i(k)$, $\hat{W}_i(k)$ and $\hat{H}_i(k)$ to simplify the notation. μ_i and λ_i are Lagrange multiplier vectors which have been introduced to take into account the equality constraints.

Since the equality constraints are from the system model, we assume the equality constraints are independent from each other. The objective function $\mathcal{F}_i$ is a quadratic function of $\hat{X}_i$ and $\hat{W}_i$, so it is continuous differentiable. Therefore, the optimal solution must satisfy the following stationary conditions of the Lagrangian, i.e. $i = 1$ to N

$$\frac{\partial \mathcal{L}}{\partial \hat{X}_i} = \frac{\partial \mathcal{F}_i}{\partial \hat{X}_i} + G_{A_{ii}}^T\mu_i + \sum_{j=1, j\neq i}^{N} G_{A_{ji}}^T\lambda_j = 0 \tag{3.3a}$$

$$\frac{\partial \mathcal{L}}{\partial \hat{W}_i} = \frac{\partial \mathcal{F}_i}{\partial \hat{W}_i} + G_{B_{ii}}^T\mu_i = 0 \tag{3.3b}$$

$$\frac{\partial \mathcal{L}}{\partial \hat{H}_i} = -\mu_i + \lambda_i = 0 \tag{3.3c}$$

$$\frac{\partial \mathcal{L}}{\partial \mu_i} = G_{A_{ii}}\hat{X}_i + G_{B_{ii}}\hat{W}_i - \hat{H}_i = 0 \tag{3.3d}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \hat{H}_i + \sum_{j=1, j\neq i}^{N} G_{A_{ij}}\hat{X}_j = 0 \tag{3.3e}$$

In the following, we show how the above centralized optimization problem can be decomposed into a few subproblems and be solved using a coordinated algorithm. A decomposition method can be used to decompose the centralized problem (3.2) into

N subproblems $\mathcal{J}_i(\alpha)$ $(i = 1, ..., N)$, where α denotes the variable in the subproblems. The key steps of decomposition methods include:

- The definition and the design of corresponding subproblems $\mathcal{J}_i(\alpha)$;
- Design of a coordination algorithm to ensure the solutions of the subproblems converge to the centralized MHE.

The Lagrange equation of centralized MHE in (3.2) can be rewritten as:

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{N} \mathcal{L}_i \\ &= \sum_{i=1}^{N} \{\mathcal{F}_i(\hat{X}_i, \hat{W}_i) + \mu_i^T (G_{A_{ii}} \hat{X}_i + G_{B_{ii}} \hat{W}_i - \hat{H}_i) + \lambda_i^T \hat{H}_i + \sum_{j=1, j\neq i}^{N} \lambda_j^T G_{A_{ji}} \hat{X}_i\}\end{aligned} \tag{3.4}$$

The Lagrange function in (3.4) takes a separable form and each subproblem can be defined by the Lagrangian $\mathcal{L}_i$ which is associated with it. The corresponding subproblem i can be defined as:

$$\begin{aligned}\min_{\hat{X}_i, \hat{W}_i, \hat{H}_i} \mathcal{J}_i &= \mathcal{F}_i(\hat{X}_i, \hat{W}_i) + \lambda_i^T \hat{H}_i + \sum_{j=1, j\neq i}^{N} \lambda_j^T G_{A_{ji}} \hat{X}_i \\ s.t.\;\; & G_{A_{ii}} \hat{X}_i + G_{B_{ii}} \hat{W}_i = \hat{H}_i\end{aligned} \tag{3.5}$$

where λ_i and λ_j will be determined by the coordinator. The part $(\lambda_i^T \hat{H}_i + \sum_{j=1, j\neq i}^{N} \lambda_j^T G_{A_{ji}} \hat{X}_i)$ is called the coordinating term denoted by $\{CoT\}_i$. $\{CoT\}_i$ links the local estimator to the coordinator.

The subproblem (3.5) can be written in the following form:

$$\begin{aligned}\min_{\hat{X}_i, \hat{W}_i, \hat{H}_i} \mathcal{J}_i = \frac{1}{2} \begin{bmatrix} \hat{X}_i(k)^T & \hat{W}_i(k)^T & \hat{H}_i(k)^T \end{bmatrix}^T & \begin{bmatrix} \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \mathbf{C}_{ii} & 0 & 0 \\ 0 & \boldsymbol{Q}_i^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \\ \hat{H}_i(k) \end{bmatrix} \\ + \{\begin{bmatrix} -\boldsymbol{Y}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} & 0 & 0 \end{bmatrix} + \boldsymbol{p}^T \varphi_i\} & \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \\ \hat{H}_i(k) \end{bmatrix} \\ = \frac{1}{2} Z_i(k)^T \Xi_i Z_i(k)^T + \{\chi_i^T + {\boldsymbol{p}^{(s)}}^T \varphi_i\} Z_i(k) & \\ s.t.\;\; F_i Z_i(k) = 0 &\end{aligned} \tag{3.6}$$

where $\hat{X}_i(k) = [\hat{\boldsymbol{x}}_i(k-Hp+1)^T, \; \cdots, \; \hat{\boldsymbol{x}}_i(k)^T]^T$, $\hat{W}_i(k) = [\hat{\boldsymbol{w}}_i(k-Hp+1)^T, \; \cdots, \; \hat{\boldsymbol{w}}_i(k-1|k)^T]^T$, $\hat{H}_i(k) = [\hat{\boldsymbol{h}}_i(k-Hp+1)^T, \; \cdots, \; \hat{\boldsymbol{h}}_i(k)^T]^T$, $Z_i(k) = [\hat{X}_i(k)^T, \hat{W}_i(k)^T, \hat{H}_i(k)^T]^T$. $\boldsymbol{p}^{(s)} = [\lambda_1^T, \; \lambda_2^T, \; \cdots, \; \lambda_N^T]^T$ is a vector with dimension $(H_p-1)n \times 1$ defined by the coordinator and transmitted to the local estimators from the coordinator, and the superscript 's' indicates that the price vector is updated iteratively by the coordinator, and n is the total number of states. The local variable $\hat{H}_i(k)$ contains the estimated interactions and is determined by the local estimators. $\hat{W}_i(k)$ is estimated process noise, as part of decision variables. The constant matrix φ_i involves the interaction model. In the equality constraint, F_i is defined as:

$$F_i = \left[\begin{array}{ccccc:cccc:cccc} -A_{ii} & I & 0 & \cdots & 0 & -I & 0 & \cdots & 0 & -I & 0 & \cdots & 0 \\ 0 & -A_{ii} & I & \cdots & 0 & 0 & -I & \cdots & 0 & 0 & -I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -A_{ii} & I & 0 & 0 & \cdots & -I & 0 & 0 & \cdots & -I \end{array}\right] = \begin{bmatrix} G_{A_{ii}} & G_{B_{ii}} & -I \end{bmatrix} \tag{3.7}$$

where F_i is a $(H_p-1)n_i \times (3H_p-2)n_i$ matrix, n_i is the number of states in subsystem i.

In 3.6, φ_i has the following form:

$$\varphi_i = \begin{bmatrix} G_{A_{1i}} & G_{B_{1i}} & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & I \\ \vdots & \vdots & \vdots \\ G_{A_{Ni}} & G_{B_{Ni}} & 0 \end{bmatrix} \begin{array}{l} \\ \\ \leftarrow i^{th} \text{ block} \\ \\ \\ \end{array} \tag{3.8}$$

where $G_{A_{ij}}$ and $G_{B_{ij}}$ are defined in (2.15) and (2.17) respectively.

The measurement matrix $\boldsymbol{Y}_i$ is defined as follows:

$$\boldsymbol{Y}_i = \begin{bmatrix} y_i(k-H_p+1) \\ y_i(k-H_p+2) \\ \vdots \\ y_i(k-1) \\ y_i(k) \end{bmatrix}$$

It should be noted that:

$$\sum_{i=1}^{N} \varphi_i \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \\ \hat{H}_i(k) \end{bmatrix} = \begin{bmatrix} \hat{h}_1(k) - \sum_{j=2}^{N} A_{1j}\hat{x}_j(k) \\ \hat{h}_2(k) - \sum_{j=1, j\neq 2}^{N} A_{2j}\hat{x}_j(k) \\ \vdots \\ \hat{h}_N(k) - \sum_{j=1}^{N-1} A_{Nj}\hat{x}_j(k) \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = \Delta E(k) \tag{3.9}$$

When $\Delta E(k) = 0$, the summation of the subproblems defined in (3.6) equals to the centralized MHE and the solution of the coordinated scheme converges to the centralized solution. If $\Delta E(k) \neq 0$, the interconnection constraints in the original centralized MHE are not satisfied, and the coordinator needs to communicate with local MHEs to continue the process until $\Delta E(k) = 0$. The information flow for this case is shown in the Figure 3.1. A local MHE sends estimated variables $\hat{X}_i$, $\hat{W}_i$ and $\hat{H}_i$ to the coordinator. The coordinator checks the stopping criterion. If the stopping criterion is satisfied, the iteration stops; otherwise the coordinator calculates the price vector $\boldsymbol{p}$ and sends it to the local MHEs to repeat the iteration.

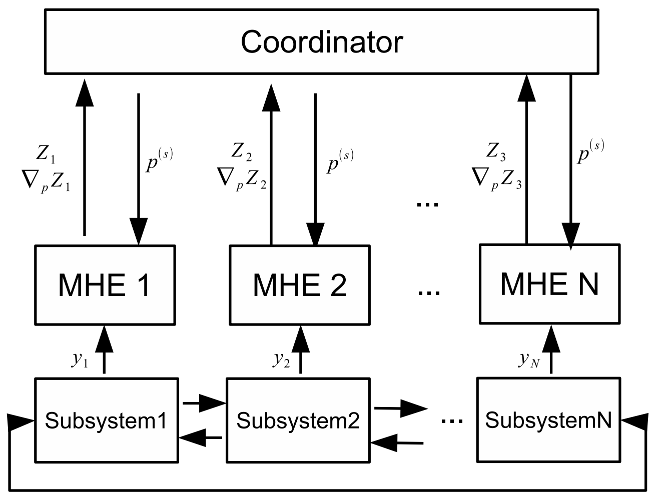

Figure 3.1: Information transfer in the price-driven CDMHE with N subsystems

In the coordinated structure, the lower layer's task is to solve equations (3.3a), (3.3b), (3.3c) and (3.3d) for a given $\boldsymbol{p}$. At the upper coordinator layer, (3.3e) should be used to update $\boldsymbol{p}$. The basis of the approach is that it is possible to convert the original minimization problem into a simpler maximization problem. The original minimization problem is the centralized MHE problem (3.1). Therefore, let us define the *dual optimization problem*.[1]

The aggregated price-driven CDMHE problem can be expressed as:

$$\min_{\hat{X},\hat{W},\hat{H}} \mathcal{J} = \sum_{i}^{N} \mathcal{J}_i(\hat{X}_i, \hat{W}_i, \hat{H}_i, \boldsymbol{p}) \tag{3.10a}$$

$$s.t: \;\; F_i Z_i = 0, \;\; i = 1, ..., N. \tag{3.10b}$$

[1]Details on duality can be found in many optimization textbooks, e.g. Boyd and Vandenberghe, 2004 and Chong and Zak, 2013.

Base on (3.4) and (3.10), the *Lagrange dual optimization problem* (Boyd and Vandenberghe, 2004) $\phi(\boldsymbol{p})$ is presented as:

$$\phi(\boldsymbol{p}) = \inf_{\hat{X},\hat{W},\hat{H}} \{\mathcal{J}(\hat{X}, \hat{W}, \hat{H}, \boldsymbol{p}) | s.t: \ F_i Z_i = 0, \ i = 1, ..., N.\} \tag{3.11}$$

Let the optimal value of the centralized MHE(3.1) be $\mathcal{R}^*$, which is also the optimal value of the price-driven CDMHE(3.10) when the algorithm converges. In (3.11), according to the property of duality in Boyd and Vandenberghe, 2004, for every $\boldsymbol{p}$ the Lagrange dual function gives is a lower bound of on the optimal value $\mathcal{R}^*$ of the optimization problem (3.10). Thus we have a lower bound that depends on the value of $\boldsymbol{p}$:

$$\phi(\boldsymbol{p}) \leq \mathcal{R}^* \tag{3.12}$$

In order to get the best lower bound from the Lagrange dual function (3.11), the following optimization problem is presented:

$$\max_{\boldsymbol{p}} \ \phi(\boldsymbol{p}) \tag{3.13}$$

where from (3.5) we can see that the price vector $\boldsymbol{p}$ in nature is the Lagrange multiplier that associated with interaction equalities (3.1c), thus the dual optimization problem (3.13) is an unconstrained optimization problem.

Denoting the optimal value of the dual optimization problem (3.13) as $\mathcal{D}^*$, from (3.12), the following conclusion can be made:

$$\mathcal{D}^* \leq \mathcal{R}^* \tag{3.14}$$

This property is called *weak duality*. The difference $\mathcal{R}^* - \mathcal{D}^*$ is referred as the *optimal duality gap* of the original problem, since it gives the gap between the optimal value of the primal problem (3.1) and the best lower bound that can be obtained from the Lagrange dual optimization problem (3.13) (Boyd and Vandenberghe, 2004).

If the equality

$$\mathcal{D}^* = \mathcal{R}^* \tag{3.15}$$

holds, then the optimal duality gap is zero, which means that the optimal value of the dual optimization problem equals to the optimal value of the primal problem. In this case, we say *strong duality* holds. The strong duality only holds when the primal problem is convex and Slater's condition holds. Slater's condition requires the primal problem has strictly feasible points.

In order to reach the optimal value of the primal optimization problem by solving the dual optimization problem, the strong duality should hold. The primal problem is the centralized MHE problem, which has quadratic function and linear constraints, and also it is assumed to have at least one feasible point. Thus, the strong duality holds which implies that the optimal value of the dual optimization problem equals to the primal optimization problem. Therefore, we can solve the dual problem in the coordinator to update the price vector $\boldsymbol{p}$.

The Lagrangian function of subsystem i (3.6) is:

$$\mathcal{L}_i = \frac{1}{2} Z_i^T \Xi_i Z_i^T + \{\chi_i^T + \boldsymbol{p}^{(s)^T} \varphi_i\} Z_i + \lambda_i^T F_i Z_i \tag{3.16}$$

The aggregated Lagrange function of the price-driven CDMHE becomes:

$$\mathcal{L}(\boldsymbol{p}, Z) = \frac{1}{2} Z^T \Xi Z^T + \{\chi^T + \boldsymbol{p}^{(s)^T} \varphi\} Z + \lambda^T F Z \tag{3.17}$$

where $Z = [Z_1^T,\ Z_2^T,\ \cdots,\ Z_N^T]^T$, $\Xi = diag(\Xi_1,\ \Xi_2,\ \cdots,\ \Xi_N)$, $\chi^T = [\chi_1^T,\ \chi_2^T,\ \cdots,\ \chi_N^T]$, $\varphi = [\varphi_1,\ \varphi_2,\ \cdots,\ \varphi_N]$, $\lambda^T = [\lambda_1^T,\ \lambda_2^T,\ \cdots,\ \lambda_N^T]$ and $F = [F_1,\ F_2,\ \cdots,\ F_N]$.

When the lower level's task is done(i.e., the subproblems are solved) , we have:

$$\frac{\partial \mathcal{L}_i}{\partial Z_i^*} = 0 \quad \text{for } i = 1,\ 2, \cdots,\ N \tag{3.18}$$

$$\frac{\partial \mathcal{L}_i}{\partial \lambda_i^*} = 0 \quad \text{for } i = 1,\ 2, \cdots,\ N \tag{3.19}$$

where Z_i^* and λ_i^* are the optimal solution of i^{th} local MHE, therefore we can get:

$$\frac{\partial \mathcal{L}}{\partial Z^*} = \Xi Z^* + \chi + \varphi^T \boldsymbol{p} + F^T \lambda^* = 0 \tag{3.20}$$

Local MHEs send the optimal Z_i^* to the coordinator. In the coordinator, the first order derivative of $\mathcal{L}$ with respect to the price vector $\boldsymbol{p}$ becomes:

$$\frac{d\mathcal{L}}{d\boldsymbol{p}} = (\Xi Z^* + \chi + \varphi^T \boldsymbol{p} + F^T \lambda^*) \frac{dZ}{d\boldsymbol{p}} + \varphi Z^* = \varphi Z^* \tag{3.21}$$

The optimal solution Z^* are sent to the coordinator as $Z^{(s)}$ to update the price vector $\boldsymbol{p}$.

Let us denote that:

$$\Delta E(k)^{(s)} = \frac{d\mathcal{L}}{d\boldsymbol{p}} = \varphi Z^{(s)} \tag{3.22}$$

In the coordinator part, we can use gradient based methods [2] to update the price vector $\boldsymbol{p}$. One efficient method is Newton's method. In this work, Newton's method is chosen to update the price vector $\boldsymbol{p}$, and the update equation at sampling time k is:

$$\boldsymbol{p}^{(s+1)}(k) = \boldsymbol{p}^{(s)}(k) - \alpha^{(s)}[H^{(s)}(k)]^{-1}\Delta E(k)^{(s)} \tag{3.23}$$

where $\alpha^{(s)}$[3] is the step size of the Newton's method at iteration s which is picked as 1 in usual if a full Newton step is taken. $H^{(s)}(k)$ is the Hessian matrix which need to be calculated. Since the dual problem is maximization problem, the Hessian matrix should be negative definite.

The Hessian matrix $H^{(s)}(k)$ is calculated as:

$$H^{(s)} = \frac{d\Delta E(k)^{(s)}}{d\boldsymbol{p}} = \varphi\frac{dZ^{(s)}}{d\boldsymbol{p}} \tag{3.24}$$

Since $\varphi Z^{(s)} = \sum_{i=1}^{N}\varphi_i Z_i^{(s)}$, equation (3.24) can be rewritten as:

$$H^{(s)} = \sum_{i=1}^{N}\varphi_i\frac{dZ_i^{(s)}}{d\boldsymbol{p}} = \sum_{i=1}^{N}\varphi_i\nabla_p Z_i^{(s)} \tag{3.25}$$

After solving the local problem (3.6), we can get solutions $\hat{X}^*_{i(s)}(k)$, $\hat{W}^*_{i(s)}(k)$ and $\hat{H}^*_{i(s)}(k)$, and send them to the coordinator. In the coordinator the states $Z_i^{(s)}$ can be obtained as: $Z_i^{(s)} = Z_i^* = [\hat{X}^*_{i(s)}(k)^T,\ \hat{W}^*_{i(s)}(k)^T,\ \hat{H}^*_{i(s)}(k)^T]^T$. The coordinator also needs the sensitivity matrix, which is defined as:

$$\triangledown_{\boldsymbol{p}} Z_i = \frac{dZ_i}{d\boldsymbol{p}} \tag{3.26}$$

According to (Cheng, 2007) and (Marcos., 2011), the way to calculate the sensitivity matrix is expressed in following procedure [4].

The condition for Z_i^* to be the optimal solution of the i^{th} local MHE can be obtained as follows:

$$\alpha_i = \frac{\partial\mathcal{L}_i}{\partial Z_i} = \Xi_i Z_i^* + \chi_i + \varphi_i^T p + F_i^T\lambda_i^* = 0 \tag{3.27a}$$

[2] Gradient method, conjugate gradient methods, Newton's method and quasi-newton methods can be used.

[3] α is picked as 1 for the unconstrained case

[4] The inequality conditions are not considered here.

$$\beta_i = F_i Z_i^* = 0 \tag{3.27b}$$

The sensitivity matrix $\nabla_p Z_i^{(s)}$ can be obtained by evaluating the gradient of Z_i with respect to $\boldsymbol{p}$:

$$\begin{cases} \dfrac{\partial \alpha_i}{\partial \boldsymbol{p}} = \Xi_i \nabla_p Z_i^* + \varphi_i^T + F_i^T \nabla_p \lambda_i^* = 0 \\ \dfrac{\partial \beta_i}{\partial \boldsymbol{p}} = F_i \nabla_p Z_i^* = 0 \end{cases} \tag{3.28}$$

Equations (3.28) can be written in matrix form as follows:

$$\begin{bmatrix} \dfrac{\partial \alpha_i}{\partial \boldsymbol{p}} \\ \dfrac{\partial \beta_i}{\partial \boldsymbol{p}} \end{bmatrix} = \begin{bmatrix} \Xi_i & F_i^T \\ F_i & 0 \end{bmatrix} \begin{bmatrix} \nabla_p Z_i^* \\ \nabla_p \lambda_i^* \end{bmatrix} = \begin{bmatrix} -\varphi_i^T \\ 0 \end{bmatrix} \tag{3.29}$$

In order to have a unique solution to the above equation, the following matrix should be invertible:

$$\Lambda_i = \begin{bmatrix} \Xi_i & F_i^T \\ F_i & 0 \end{bmatrix} \tag{3.30}$$

If Λ_i is invertible, $\nabla_{\boldsymbol{p}} Z_i^*$ can be picked as the first $(3H_p - 1)n_i$ rows of Λ_i.

In summary, for a given price vector $\boldsymbol{p}^{(s)}$, each local MHE solves the local optimization problem (3.6) to get Z_i^* (for $i = 1, 2, \cdots, N$), calculates sensitivity matrix $\nabla_{\boldsymbol{p}} Z_i^*$, and then sends them to the coordinator. The coordinator uses Z^* as $Z^{(s)}$ to check the stopping criterion. If the interaction error $\| \Delta E(k)^{(s)} \| \leq \epsilon$, the iteration stops; otherwise, the Hessian matrix $H^{(s)}$ is calculated according to equation (3.25). The coordinator updates price vector $\boldsymbol{p}$ according to (3.23). Iteration counter increases by one (i.e., $s \leftarrow s + 1$) and repeats the process. The detail price-driven CDMHE algorithm is described in the following table:

3.2 An Improved Price-driven CDMHE

In Section 3.1, we proposed a price-driven CDMHE algorithm without consideration of inequality constraints. Since the problem is unconstrained, we can find the analytical solution of the subsystem optimization problems. The first order optimality condition

Algorithm 1 Implementation of Price-driven Coordinated Distributed MHE

Initialization

Coordinator: Set iteration counter $s = 0$. The price vector $\boldsymbol{p}^{(0)}$ is arbitrarily determined.

repeat

Coordinator: $\boldsymbol{p}^{(s)}$ is sent to local MHE estimators.

Local Estimators: Local MHE problems (3.6) are solved, and the optimal solutions $\hat{X}^*_{i(s)}(k)$, $\hat{W}^*_{i(s)}(k)$ and $\hat{H}^*_{i(s)}(k)$ are obtained. Construct $Z^*_i = [\hat{X}^*_{i(s)}(k)^T, \hat{W}^*_{i(s)}(k)^T, \hat{H}^*_{i(s)}(k)^T]^T$. $\nabla_{\boldsymbol{p}} Z^*_i$ is calculated according to equation (3.29). $Z^*_i(k)$ and $\nabla_{\boldsymbol{p}} Z^*_i$ (for $i = 1, 2, \cdots, N$) are sent to the coordinator.

Coordinator: $s \leftarrow s+1$. Coordinator receives $Z^*_i(k)$ and $\nabla_{\boldsymbol{p}} Z^*_i$, and treats $Z^*_i(k)$ as $Z^{(s)}_i(k)$. The Hessian matrix $H^{(s)}$ is calculated according to (3.25). Step-size α is chosen/calculated. Price vector $\boldsymbol{p}$ is updated according to (3.23).

until stopping criterion $\parallel \Delta E(k)^{(s)} \parallel \leq \epsilon$ is satisfied.

for a variable Z^*_i to be the optimal solution of problem (3.6) is listed as following:

$$\frac{\partial \mathcal{L}_i}{\partial Z_i} = \Xi_i Z^*_i + \chi_i + \varphi_i^T \boldsymbol{p}^{(s)} + F_i^T \lambda^*_i = 0 \tag{3.31a}$$

$$\frac{\partial \mathcal{L}_i}{\partial \lambda_i} = F_i Z^*_i = 0 \tag{3.31b}$$

If a unique solution exists for the of the above equation (3.31):

$$\begin{bmatrix} Z^*_i \\ \lambda^*_i \end{bmatrix} = \begin{bmatrix} \Xi_i & F_i^T \\ F_i & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\varphi_i{}^T \boldsymbol{p}^{(s)} - \chi_i \\ 0 \end{bmatrix} \tag{3.32}$$

In order to get the unique solution and use the Newton's method to update the price vector, we need to prove the invertibility of Λ_i. However, the invertibility of Λ_i can only be guaranteed when the matrix C_{ii} is full rank. A detail proof can be found in the Appendix A.1. For the estimation problem, the C_{ii} is not full rank in usual. Due to this, we cannot get unique solutions from solving the local optimization problem (3.6) and the Newton's method can not be used to update the price vector [5].

One way to make the matrix Λ_i invertible is to add an additional term in the cost function of subsystem i (3.6). In this work, we propose to add $\frac{1}{2} \parallel \hat{H}_i(k) \parallel_{\boldsymbol{D}^{-1}}$.

[5] Other gradient based method like gradient method can be used to update the price vector by

$$\boldsymbol{p}^{(s+1)} = \boldsymbol{p}^{(s)} - \mathcal{K}\frac{d\mathcal{L}}{d\boldsymbol{p}} = \boldsymbol{p}^{(s)} - \mathcal{K}\sum_{i=1}^{N} \varphi_i Z_i^{(s)} \tag{3.33}$$

where $\mathcal{K}$ is the step size.

Therefore the improved design of local DMHE i becomes:

$$
\begin{aligned}
\min_{\hat{X}_i,\hat{W}_i,\hat{H}_i} \mathcal{J}_i &= \frac{1}{2}\begin{bmatrix}\hat{X}_i(k)^T & \hat{W}_i(k)^T & \hat{H}_i(k)^T\end{bmatrix}^T \begin{bmatrix} \boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\mathbf{C}_{ii} & 0 & 0 \\ 0 & \boldsymbol{Q}_i^{-1} & 0 \\ 0 & 0 & \boldsymbol{D}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\\ \hat{H}_i(k)\end{bmatrix} \\
&+ \{\begin{bmatrix}-\boldsymbol{Y}_i^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & 0 & 0\end{bmatrix} + \boldsymbol{p}^T\varphi_i\}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\\ \hat{H}_i(k)\end{bmatrix} \\
&= \frac{1}{2}Z_i(k)^T\Xi_i^* Z_i(k)^T + \{\chi_i^T + \boldsymbol{p}^{(s)^T}\varphi_i\}Z_i(k) \\
s.t. \quad & F_i Z_i(k) = 0
\end{aligned} \tag{3.34}
$$

where Ξ_i^* is used to differentiate from Ξ_i in (3.6), $\boldsymbol{D}_i^{-1}$ is the weighting matrix of $\hat{H}_i(k)$. The addition of the term $\hat{H}_i(k)^T\boldsymbol{D}_i^{-1}\hat{H}_i(k)$ ensures that problem (3.34) has a unique solution. $\boldsymbol{D}_i^{-1}$ needs to satisfy certain conditions which can be found in Appendix A.2.

We note that the improved design of the local MHE includes the interaction in the cost function which is added purely to ensure the uniqueness of the solutions of the local MHEs.

3.3 Convergence Properties of the Improved Price-driven CDMHE

In this section, we show that the improved price-driven CDMHE can converge in two iterations at each sampling time.

As shown in equation (3.31), the first order optimal condition for the modified optimization problem of local MHE i at s^{th} communication cycle can be written as:

$$
\frac{\partial \mathcal{L}_i}{\partial Z_i} = \begin{bmatrix} \boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\mathbf{C}_{ii} & 0 & 0 \\ 0 & \boldsymbol{Q}_i^{-1} & 0 \\ 0 & 0 & \boldsymbol{D}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i^*(k)\\ \hat{W}_i^*(k)\\ \hat{H}_i^*(k)\end{bmatrix} + \begin{bmatrix}-\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{Y}_i \\ 0 \\ 0\end{bmatrix} + \begin{bmatrix}\varphi_{iA}^T \\ \varphi_{iB}^T \\ \varphi_{iC}^T\end{bmatrix}\boldsymbol{p}^{(s)} + \begin{bmatrix}G_{A_{ii}}^T \\ -I \\ -I\end{bmatrix}\lambda_i^* = 0 \tag{3.35a}
$$

$$
\frac{\partial \mathcal{L}_i}{\partial \lambda_i} = \begin{bmatrix}G_{A_{ii}} & -I & -I\end{bmatrix}\begin{bmatrix}\hat{X}_i^*(k)\\ \hat{W}_i^*(k)\\ \hat{H}_i^*(k)\end{bmatrix} = 0 \tag{3.35b}
$$

where $\varphi = [\varphi_{iA},\ \varphi_{iB},\ \varphi_{iC}]$.

Equation (3.35) can be written into the following form:

$$\begin{cases} \boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii}\hat{X}_i^*(k)+\varphi_{iA}^T\boldsymbol{p}^{(s)}+G_{A_{ii}}^T\lambda_i^*=\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{Y}_i \\ \boldsymbol{Q}_i^{-1}\hat{W}_i^*(k)+\varphi_{iB}^T\boldsymbol{p}^{(s)}-\lambda_i^*=0 \\ \boldsymbol{D}_i^{-1}\hat{H}_i^*(k)+\varphi_{iC}^T\boldsymbol{p}^{(s)}-\lambda_i^*=0 \\ G_{A_{ii}}\hat{X}_i^*(k)-\hat{W}_i^*(k)-\hat{H}_i^*(k)=0 \end{cases} \tag{3.36}$$

The optimality conditions (3.36) for subsystems can be aggregated from $i=1$ to N as:

$$\begin{cases} \mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C}\hat{X}_{MHE}+\varphi_A^T\boldsymbol{p}^{(s)}+\bar{G}_A^T\lambda_{MHE}=\mathbb{C}^T\boldsymbol{R}^{-1}\boldsymbol{Y} \\ \boldsymbol{Q}^{-1}\hat{W}_{MHE}+\varphi_B^T\boldsymbol{p}^{(s)}-\lambda_{MHE}=0 \\ \boldsymbol{D}^{-1}\hat{H}_{MHE}+\varphi_C^T\boldsymbol{p}^{(s)}-\lambda_{MHE}=0 \\ \bar{G}_A\hat{X}_{MHE}-\hat{W}_{MHE}-\hat{H}_{MHE}=0 \end{cases} \tag{3.37}$$

where $\boldsymbol{D}=diag(\boldsymbol{Q}_1,\ \ldots,\boldsymbol{Q}_N)$ is the aggregated form of $\boldsymbol{Q}_i$, $\mathbb{C}^T$ and $\boldsymbol{R}^{-1}$ are defined in 2.10) and (2.11), $\bar{G}_A=diag(G_{A_{11}},\ \ldots,\ G_{A_{NN}})$, $\varphi_A=[\varphi_{A_1},\ \varphi_{A_2},\ \ldots,\ \varphi_{A_N}]$, $\varphi_B=[\varphi_{B_1},\ \varphi_{B_2},\ \ldots,\ \varphi_{B_N}]$ and $\varphi_C=[\varphi_{C_1},\ \varphi_{C_2},\ \ldots,\ \varphi_{C_N}]$. In order to simplify the notation, the sampling time k is not shown in the estimated variables. $\hat{X}_{MHE}$, $\hat{W}_{MHE}$ and $\hat{H}_{MHE}$ are the aggregated forms of the optimal solution of subsystems.

From the definition of φ_i in (3.8) and the definition of $G_{B_{ij}}$ in (2.17), it can be concluded that $\varphi_B=0$ and $\varphi_C=-I$. Substituting these into the equation (3.37) and solve the equations. From the second equation in (3.37), we can get:

$$\lambda_{MHE}=\boldsymbol{Q}^{-1}\hat{W}_{MHE} \tag{3.38}$$

By combining equation (3.38) with the third and fourth equations in (3.37), we can get that:

$$\hat{W}_{MHE}=(\boldsymbol{D}\boldsymbol{Q}^{-1}+I)^{-1}(\bar{G}_A\hat{X}_{MHE}-\boldsymbol{D}\boldsymbol{p}^{(s)}) \tag{3.39}$$

Substituting equations (3.39) and (3.38) into the first equation in (3.37), $\hat{X}_{MHE}$ can be obtained as:

$$\hat{X}_{MHE}=(\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C}+\bar{G}_A\boldsymbol{Q}^{-1}(\boldsymbol{D}\boldsymbol{Q}^{-1}+I)^{-1}\bar{G}_A)^{-1}\{[(\boldsymbol{D}\boldsymbol{Q}^{-1}+I)^{-1}D-\varphi_A^T]\boldsymbol{p}^{(s)}+\mathbb{C}^T\boldsymbol{R}^{-1}\boldsymbol{Y}\} \tag{3.40}$$

Therefore, at s^{th} iteration, $\hat{X}_{MHE}$ and $\hat{W}_{MHE}$ can be expressed as a function of $\boldsymbol{p}^s$ as:

$$\hat{X}_{MHE}=c_1\boldsymbol{p}^{(s)}+c_2 \tag{3.41a}$$

$$\hat{W}_{MHE} = c_3\boldsymbol{p}^{(s)} + c_4 \tag{3.41b}$$

with

$$c_1 = (\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C} + \bar{G}_A\boldsymbol{Q}^{-1}(\boldsymbol{D}\boldsymbol{Q}^{-1} + I)^{-1}\bar{G}_A)^{-1}[(\boldsymbol{D}\boldsymbol{Q}^{-1} + I)^{-1}D - \varphi_A^T] \tag{3.42a}$$

$$c_2 = (\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C} + \bar{G}_A\boldsymbol{Q}^{-1}(\boldsymbol{D}\boldsymbol{Q}^{-1} + I)^{-1}\bar{G}_A)^{-1}\mathbb{C}^T\boldsymbol{R}^{-1}\boldsymbol{Y} \tag{3.42b}$$

$$c_3 = (\boldsymbol{D}\boldsymbol{Q}^{-1} + I)^{-1}(\bar{G}_Ac_1 - \boldsymbol{D}) \tag{3.42c}$$

$$c_4 = (\boldsymbol{D}\boldsymbol{Q}^{-1} + I)^{-1}\bar{G}_Ac_2 \tag{3.42d}$$

The expression of c_1 shows that c_1 is only dependent on system matrices $A,\ B,\ C$ and weighting matrices $\boldsymbol{Q}^{-1},\ \boldsymbol{R}^{-1},\ \boldsymbol{D}^{-1}$ and horizon H_p, while the matrix c_2 is dependent on $A,\ B,\ C,\ \boldsymbol{Q}^{-1},\ \boldsymbol{R}^{-1},\ \boldsymbol{D}^{-1}$, horizon H_p and output $\boldsymbol{Y}$. $A,\ B,\ C$ are system matrices which represent system properties and remain unchanged for a linear time-invariant system. $\boldsymbol{Q}^{-1},\ \boldsymbol{R}^{-1},\ \boldsymbol{D}^{-1}$ and H_p are from the local MHE design, which are fixed for an existing decentralized network. Output $\boldsymbol{Y}$ remains unchanged during the estimation interval. Therefore the linear coefficients c_1 and c_2 are constant matrices during the k^{th} estimation interval.

From equation (3.22), we can get:

$$\Delta E(k)^{(s)} = \varphi Z^{(s)} = \varphi_A\hat{X}_{MHE} + \varphi_B\hat{W}_{MHE} + \varphi_C\hat{H}_{MHE} \tag{3.43}$$

Since $\varphi_B = 0$ and $\varphi_C = -I$, equation (3.43) becomes:

$$\begin{aligned}\Delta E(k)^{(s)} &= \varphi_A\hat{X}_{MHE} - \hat{H}_{MHE} \\ &= \varphi_A\hat{X}_{MHE} + \bar{G}_A\hat{X}_{MHE} - \hat{W}_{MHE} \\ &= ((\varphi_A + \bar{G}_A)c_1 - c_3)\boldsymbol{p}^{(s)} + c_2 - c_4\end{aligned} \tag{3.44}$$

From equation (3.24), the Hessian matrix is calculated as:

$$H^{(s)} = \frac{d\Delta E(k)^{(s)}}{d\boldsymbol{p}^{(s)}} = (\varphi_A + \bar{G}_A)c_1 - c_3 \tag{3.45}$$

Therefore $\Delta E(k)^{(s)}$ can be expressed as:

$$\Delta E(k)^{(s)} = H^{(s)}\boldsymbol{p}^{(s)} + c_2 - c_4 \tag{3.46}$$

Substituting equation (3.45) into equation (3.23), and picking $\alpha^{(s)}$ to be 1, the following equation can be obtained:

$$\begin{aligned}\boldsymbol{p}^{(s+1)}(k) &= \boldsymbol{p}^{(s)}(k) - \alpha^{(s)}[H^{(s)}(k)]^{-1}\Delta E(k)^{(s)} \\ &= \boldsymbol{p}^{(s)}(k) - [H^{(s)}(k)]^{-1}(H^{(s)}\boldsymbol{p}^{(s)} + c_2 - c_4) \\ &= -[H^{(s)}(k)]^{-1}(c_2 - c_4) \\ &= \varsigma\end{aligned} \tag{3.47}$$

For the unconstrained price-driven CDMHE, $H^{(s)}(k)^{-1},\ c_2,\ c_4$ are constant matrices as defined in equations (3.45), (3.42b) and (3.42d). When the iteration $s \geq 2$, $\boldsymbol{p}^{(s)}(k)$ becomes a constant matrix ς. Therefore, when the iteration $s \geq 2$, $\Delta E(k)^{(s)}$ becomes:

$$\begin{aligned}\Delta E(k)^{(s)} &= H^{(s)}\boldsymbol{p}^{(s)} + c_2 - c_4 \\ &= -H^{(s)}[H^{(s)}(k)]^{-1}(c_2 - c_4) + c_2 - c_4 \\ &= 0 \qquad (s \geq 2)\end{aligned} \tag{3.48}$$

From equation (3.48), it can be concluded that the improved price-driven CDMHE algorithm converges in 2 iterations without consideration of inequality constraints.

3.4 Improved Price-driven CDMHE with Inequality Constraints

In this section, inequality constraints are added to the estimated state $\hat{\boldsymbol{x}}_i$, process noise $\hat{\boldsymbol{w}}_i$ and estimated interaction $\hat{\boldsymbol{h}}_i$. The sensitivity matrices that used to calculate the Hessian matrix are derived with inequality constraints. Inequality constraints are divided into active inequality constraints and inactive inequality constraints in each subsystem. Due to the change of the active inequality constraints, the Hessian matrix needs to be calculated in each iteration.

3.4.1 Sensitivity Matrix Calculation

For subsystem i, the local MHE design becomes:

$$\begin{aligned}\min_{\hat{X}_i,\hat{W}_i,\hat{H}_i} \mathcal{J}_i &= \frac{1}{2}Z_i(k)^T\Xi_i^* Z_i(k)^T + \{\chi_i^T + {\boldsymbol{p}^{(s)}}^T\varphi_i\}Z_i(k) \\ s.t.\ \ & F_i Z_i(k) = 0 \\ & F_i^{ineq} Z_i(k) \leq g_i^{ineq}\end{aligned} \tag{3.49}$$

where $F_i^{ineq} Z_i(k) \leq g^{ineq}$ is used to denote the constraints that added to $\hat{X}_i$, $\hat{W}_i$ and $\hat{H}_i$(i.e., $\hat{X}_i \in \mathbb{X}_i$, $\hat{W}_i \in \mathbb{W}_i$, $\hat{H}_i \in \mathbb{H}_i$). F_i^{ineq} and g^{ineq} have the following expressions:

$$F_i^{ineq} = \begin{bmatrix} I_{H_p n_i} & 0 & 0 \\ -I_{H_p n_i} & 0 & 0 \\ 0 & I_{(H_p-1)n_i} & 0 \\ 0 & -I_{(H_p-1)n_i} & 0 \\ 0 & 0 & I_{(H_p-1)n_i} \\ 0 & 0 & -I_{(H_p-1)n_i} \end{bmatrix} \tag{3.50}$$

$$g_i^{ineq} = \begin{bmatrix} X_{i_{max}} \\ -X_{i_{min}} \\ W_{i_{max}} \\ -W_{i_{min}} \\ H_{i_{max}} \\ -H_{i_{min}} \end{bmatrix} \tag{3.51}$$

The Lagrange function L_i can be formulated according to the optimization problem defined in equation (3.49) as:

$$L_i = \frac{1}{2} Z_i^T \Xi_i^* Z_i^T + \{\chi_i^T + \boldsymbol{p}^{(s)^T} \varphi_i\} Z_i + \lambda_i^T F_i Z_i + \mu_i^T (F_i^{ineq} Z_i(k) - g_i^{ineq}) \tag{3.52}$$

where the vectors λ_i and μ_i are the Lagrange multipliers that associated with the equality constraints and inequality constraints in equation (3.49).

The first order optimality conditions are:

$$\frac{\partial L_i}{\partial Z_i} = \Xi_i^* Z_i + \chi_i + \varphi_i^T \boldsymbol{p}^{(s)} + F_i^T \lambda_i + {}_{\mathcal{AC}}F_i^{ineq^T} {}_{\mathcal{AC}}\mu_i = 0 \tag{3.53a}$$

$$\frac{\partial L_i}{\partial \lambda_i} = F_i Z_i = 0 \tag{3.53b}$$

$${}_{\mathcal{AC}}F_i^{ineq} Z_i - {}_{\mathcal{AC}}g_i^{ineq} = 0 \tag{3.53c}$$

$${}_{\mathcal{IN}}F_i^{ineq} Z_i - {}_{\mathcal{IN}}g_i^{ineq} + {}_{\mathcal{IN}}\sigma_i = 0 \tag{3.53d}$$

where the subscript $\mathcal{AC}$ stands for the active inequality constraints, $\mathcal{IN}$ stands for the inactive inequality constraints. The vector σ_i represents the slack variable that associated with inactive inequality constraints. The sensitivity analysis can be conducted as follows:

$$\Xi_i^* \nabla_p Z_i + \varphi_i^T + F_i^T \nabla_p \lambda_i + {}_{\mathcal{AC}}F_i^{ineq^T} \nabla_{p\mathcal{AC}} \mu_i = 0 \tag{3.54a}$$

$$F_i \nabla_p Z_i = 0 \tag{3.54b}$$

$${}_{\mathcal{AC}}F_i^{ineq} \nabla_p Z_i = 0 \tag{3.54c}$$

$$
{}_{\mathcal{IN}}F_i^{ineq} \nabla_p Z_i + \nabla_{p\mathcal{IN}} \sigma_i = 0 \tag{3.54d}
$$

where $\nabla_p Z_i$ denotes dZ_i/dp^s, $\nabla_p \lambda_i$ denotes $d\lambda_i/dp^s$, $\nabla_{p\mathcal{AC}}\mu_i$ represents $d_{\mathcal{AC}}\mu_i/dp^s$, and $\nabla_{p\mathcal{IN}}\sigma_i$ represents $d_{\mathcal{IN}}\sigma_i/dp^s$.

Therefore, the sensitivity matrix $\nabla_p Z_i$ can be calculated by solving the equations in (3.54). Equation (3.54) can be written into the matrix form as:

$$
\Pi_i \begin{bmatrix} \nabla_p Z_i \\ \nabla_p \lambda_i \\ \nabla_{p\mathcal{AC}} \mu_i \\ \nabla_{p\mathcal{IN}} \sigma_i \end{bmatrix} = \begin{bmatrix} -\varphi_i^T \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.55}
$$

with

$$
\Pi_i = \begin{bmatrix} \Xi_i^* & F_i^T & {}_{\mathcal{AC}}F_i^{ineqT} & 0 \\ F_i & 0 & 0 & 0 \\ {}_{\mathcal{AC}}F_i^{ineq} & 0 & 0 & 0 \\ {}_{\mathcal{IN}}F_i^{ineq} & 0 & 0 & I \end{bmatrix} \tag{3.56}
$$

Hessian matrix is calculated as:

$$
H^{(s)} = \sum_{i=1}^{N} \varphi_i \frac{dZ_i^{(s)}}{d\boldsymbol{p}} = \sum_{i=1}^{N} \varphi_i \nabla_p Z_i^{(s)} \tag{3.57}
$$

Step size α in the price vector updating equation (3.23) is calculated using the method in Cheng, 2007.

3.4.2 Convergence Analysis

In this section, the convergence of the constrained modified price-driven CDMHE algorithm is proven by using the property of the gradient-based method.

Define the function $\mathcal{W}$ as following:

$$
\mathcal{W}(\boldsymbol{p}) = \mathcal{D}^* - \phi(\boldsymbol{p}) \tag{3.58}
$$

where $\mathcal{D}^*$ is the optimal value of the dual optimization problem (3.13), and $\phi(\boldsymbol{p})$ is defined in (3.11). Therefore, $\mathcal{W} \geq 0$ and equals to zero if and only if $\boldsymbol{p} = \boldsymbol{p}^*$.

The derivative of $\mathcal{W}$ with respect to the iterations s is expressed by:

$$
\begin{aligned}
\dot{\mathcal{W}} &= \frac{d\mathcal{W}}{d\boldsymbol{p}}^T \frac{d\boldsymbol{p}}{ds} \\
&= (-\frac{d\phi(\boldsymbol{p})}{d\boldsymbol{p}})^T \frac{d\boldsymbol{p}}{ds}
\end{aligned} \tag{3.59}
$$

The gradient

$$\frac{d\phi(\boldsymbol{p})}{d\boldsymbol{p}} = \varphi Z = J \tag{3.60}$$

where J is used to represent the gradient.

As discussed in the coordinator design part, the gradient-based method is used to update the price vector $\boldsymbol{p}$. The dual problem (3.13) we solve to get $\boldsymbol{p}$ is a maximization problem. For the maximization problem, the rate of change of the price vector is proportional to the gradient vector. Therefore, the following equation can be obtained:

$$\frac{d\boldsymbol{p}}{ds} \propto J = \nu J \tag{3.61}$$

where ν is a positive constant.

Therefore, substituting (3.61) and (3.60) into (3.59), the following expression can be obtained:

$$\dot{\mathcal{W}} = -\nu J^T J \tag{3.62}$$

where ν is positive, we can get $\dot{\mathcal{W}} \leq 0$. Therefore, the price-update scheme is stable and convergent. This implies $\Delta E(k)^{(s)} \to 0$ with the increase of iteration numbers s. Thus, the proposed price-driven CDMHE algorithm converges to the optimal solution of the centralized MHE.

3.5 Illustrative Example

In this section, a chemical process is used to illustrate the effectiveness and applicability of the proposed price-driven CDMHE scheme.

3.5.1 Problem Description

In this section, the proposed coordinated state estimation approach is applied to a simulated chemical process. The process contains two connected continuous stirred tank reactors (CSTR) and one flash tank separator as shown in Figure 3.2 (Liu *et al.*, 2008). As shown in Figure 3.2, pure A is fed into the first CSTR at flow rate F_{10} and temperature T_{10}. The outlet stream of CSTR 1 is fed to CSTR 2 at flow rate F_1 and temperature T_1. There is an additional stream containing pure A at flow

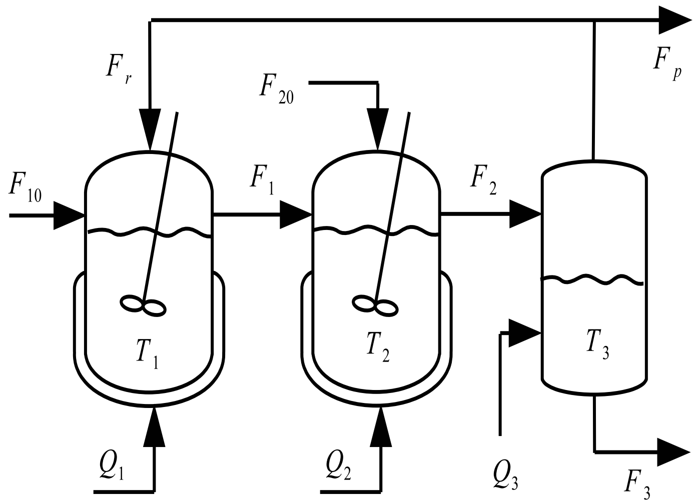

Figure 3.2: Process flow diagram of two interconnected CSTR units and one separator.

rate F_{20} and temperature T_{20} which is also fed to CSTR 2. A part of the output of CSTR 2 passes through a separator and recycled back to CSTR 1 at recycle flow rate F_r and temperature T_3. Two irreversible elementary exothermic reactions $A \to B$, $B \to C$ take place in the two reactors, where A is the reactant material, B is the desired product, and C is the undesired byproducts. Because of the non-isothermal nature of the reactions, each CSTR is equipped with a jacket to remove/provide heat. Based on standard modeling assumptions, nine ordinary differential equations can be obtained to describe the dynamics:

$$\frac{dx_{A_1}}{dt} = \frac{F_{10}}{V_1}(x_{A10} - x_{A_1}) + \frac{F_r}{V_1}(x_{Ar} - x_{A1}) - k_1 e^{\frac{-E_1}{RT_1}} x_{A_1} \tag{3.63a}$$

$$\frac{dx_{B_1}}{dt} = \frac{F_{10}}{V_1}(x_{B10} - x_{B_1}) + \frac{F_r}{V_1}(x_{Br} - x_{B1}) + k_1 e^{\frac{-E_1}{RT_1}} x_{A_1} - k_2 e^{\frac{-E_2}{RT_1}} x_{B_1} \tag{3.63b}$$

$$\frac{dT_1}{dt} = \frac{F_{10}}{V_1}(T_{10} - T_1) + \frac{F_r}{V_1}(T_3 - T_1) + \frac{-\Delta H_1}{c_p} k_1 e^{\frac{-E_1}{RT_1}} x_{A_1} + \frac{-\Delta H_2}{c_p} k_2 e^{\frac{-E_2}{RT_1}} x_{B_1} + \frac{Q_1}{\rho c_p V_1} \tag{3.63c}$$

$$\frac{dx_{A_2}}{dt} = \frac{F_1}{V_2}(x_{A_1} - x_{A_2}) + \frac{F_{20}}{V_2}(x_{A20} - x_{A_2}) - k_1 e^{\frac{-E_1}{RT_2}} x_{A_2} \tag{3.63d}$$

$$\frac{dx_{B_2}}{dt} = \frac{F_1}{V_2}(x_{B_1} - x_{B_2}) + \frac{F_{20}}{V_2}(x_{B20} - x_{B_2}) + k_1 e^{\frac{-E_1}{RT_2}} x_{A_2} - k_2 e^{\frac{-E_2}{RT_2}} x_{B_2} \tag{3.63e}$$

$$\frac{dT_2}{dt} = \frac{F_1}{V_2}(T_1 - T_2) + \frac{F_{20}}{V_2}(T_{20} - T_2) + \frac{Q_2}{\rho c_p V_2} + \frac{-\Delta H_1}{c_p} k_1 e^{\frac{-E_1}{RT_2}} x_{A_2} + \frac{-\Delta H_2}{c_p} k_2 e^{\frac{-E_2}{RT_2}} x_{B_2} \tag{3.63f}$$

Table 3.1: Process variables for the reactors and separator

x_{A1}, x_{A2}, x_{A3}	mass fractions of A in reactors 1,2,3
x_{B1}, x_{B2}, x_{B3}	mass fractions of B in reactors 1,2,3
x_{C1}, x_{C2}, x_{C3}	mass fractions of A in reactors 1,2,3
x_{Ar}, x_{Br}, x_{Cr}	mas fractions of A,B,C in the recycel stream
T_1, T_2, T_3	temperatures in reactors 1,2,3
T_{10}, T_{20}	feed stream temperatures to reactors 1 and 2
F_1, F_2	effluent flow rates from reactors 1 and 2
F_10, F_20	steady-state feed stream flow rates to reactors 1 and 2
F_r, F_p	flow rates of the recycle and purge streams
V_1, V_2, V_3	volumes of reactors 1, 2, 3
E_1,E_2	activation energy for reactions $A \to B$ and $B \to C$
k_1, k_2	pre-exponential values for reactions $A \to B$ and $B \to C$
$\Delta H_1, \Delta H_2$	heats of reaction for reactions $A \to B$ and $B \to C$
$\Delta H_{vap1}, \Delta H_{vap2}, \Delta H_{vap3}$	evaporating enthalpies for A,B,C
$\alpha_A, \alpha_B, \alpha_C$	relative volatilities of A,B,C
Q_1, Q_2, Q_3	heat inputs/removals into/from reactors 1,2,3
ρ, c_p, R	solution density, heat capacity and gas constant

$$\frac{dx_{A_3}}{dt} = \frac{F_2}{V_3}(x_{A_2} - x_{A_3}) - \frac{(F_r + F_p)}{V_3}(x_{Ar} - x_{A_3}) \tag{3.63g}$$

$$\frac{dx_{B_3}}{dt} = \frac{F_2}{V_3}(x_{B_2} - x_{B_3}) - \frac{(F_r + F_p)}{V_3}(x_{Br} - x_{B_3}) \tag{3.63h}$$

$$\frac{dT_3}{dt} = \frac{F_2}{V_3}(T_2 - T_3) + \frac{Q_3}{\rho c_p V_3} + \frac{(F_r + F_p)}{\rho c_p V_3}(x_{Ar}\Delta H_{vap1} + x_{Br}\Delta H_{vap2} + x_{Cr}\Delta H_{vap3}) \tag{3.63i}$$

where values of the parameters are given in the Table 1. It is assumed that in the separator, the relative volatility for each of the components remains constant within the operating temperature range. Under this assumption, the algebraic equations modeling the composition of the overhead stream relative to composition of liquid in the flash tank is described as follows:

$$\begin{aligned} x_{Ar} &= \frac{\alpha_A x_{A3}}{\alpha_A x_{A3} + \alpha_B x_{B3} + \alpha_C x_{C3}} \\ x_{Br} &= \frac{\alpha_B x_{B3}}{\alpha_A x_{A3} + \alpha_B x_{B3} + \alpha_C x_{C3}} \\ x_{Cr} &= \frac{\alpha_C x_{C3}}{\alpha_A x_{A3} + \alpha_B x_{B3} + \alpha_C x_{C3}} \end{aligned} \tag{3.64}$$

The sytem is divided into three subsystems according to the three vessels in the process, the states are expressed as $x_i = [x_{Ai},\ x_{Bi},\ T_i]^T$, for $i = 1,\ 2, 3$.

Table 3.2: Process parameters for the reactors and separator

$F_{10} = 5.04m^3/h$	$k_1 = 2.77 \times 10^3 s^{-1}$
$F_{20} = 5.04m^3/h$	$k_2 = 2.6 \times 10^3 s^{-1}$
$F_r = 5.04m^3/h$	$E_1 = 5.0 \times 10^4 kJ/kmol$
$F_p = 5.04m^3/h$	$E_2 = 6.0 \times 10^4 kJ/kmol$
$V_1 = 1.0m^3$	$T_{10} = 300K$
$V_2 = 0.5m^3$	$T_{20} = 300K$
$V_3 = 1.0m^3$	$R = 8.314kJ/kmol\ K$
$\alpha_A = 3.5$	$\rho = 1000.0kg/m^3$
$\alpha_B = 1.0$	$c_p = 0.231kJ/kgK$
$\alpha_C = 0.5$	$x_{A10} = 1$
$T_{10} = 300.0K$	$x_{B10} = 0$
$T_{20} = 300.0K$	$x_{A20} = 0$
$\Delta H_{vap1} = -3.53 \times 10^4 kJ/kmol$	$x_{B20} = 0$
$\Delta H_{vap2} = -1.57 \times 10^4 kJ/kmol$	$\Delta H_1 = -6.0 \times 10^4 kJ/kmol$
$\Delta H_{vap3} = -4.068 \times 10^4 kJ/kmol$	$\Delta H_2 = -7.0 \times 10^4 kJ/kmol$

The corresponding external input to each vessels are Q_1, Q_2 and Q_3. It is assumed that the measurement states are temperatures. Choose the input heat $Q = [2.9 \times 10^6 kJ/h,\ 1.0 \times 10^6 kJ/h,\ 2.9 \times 10^6 kJ/h]^T$, the corresponding steady state is

$$x_s = \begin{bmatrix} 0.0313 & 0.2602 & 549.7686 & 0.1203 & 0.3499 & 499.4590 & 0.0348 & 0.2855 & 535.5199 \end{bmatrix}^T.$$

Linearize the nonlinear model at x_s and discretize the continuous model into discrete time model by sampling time 0.005 hr, we can get the linear discrete time model. The random disturbances added to the dynamics of the temperatures' differences are generated as normally distributed values with zero mean and standard deviation 0.01, bounded between $[-0.05,\ 0.05]$. The random disturbances added to the dynamics of the concentrations are generated as normally distributed values with zero mean and standard deviation 0.001, bounded between $[-0.005,\ 0.005]$. The concentrations are bounded between$[0,\ 1]$, the temperatures are bounded between $[500,\ 600]$. In this case, we do not add constraints on the estimated interaction The parameters used in the coordinated algorithm are in Table 4.3. For the price-driven CDMHE algorithm, the termination threshold ϵ is 0.001. The actual initial condition is

$$x_{init} = \begin{bmatrix} 0.0443 & 0.3052 & 557.2686 & 0.1413 & 0.4129 & 508.7090 & 0.0578 & 0.3325 & 542.8199 \end{bmatrix}^T.$$

Table 3.3: Parameters used in the price-driven CDMHE

	Initial Guess	Moving Horizon	Weighting Matrix
MHE 1	$\hat{x}_1(0) = [0.0508\ 0.3277\ 561.0186]^T$	10	$Q_1 = \begin{bmatrix} 0.001^2 & 0 & \\ 0 & 0.001^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}$ $R_1 = 0.01^2$ $D_1 = \begin{bmatrix} 1^2 & 0 & \\ 0 & 1^2 & 0 \\ 0 & 0 & 1^2 \end{bmatrix}$
MHE 2	$\hat{x}_2(0) = [0.1518\ 0.4444\ 513.3340]^T$	10	$Q_2 = \begin{bmatrix} 0.001^2 & 0 & \\ 0 & 0.001^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}$ $R_2 = 0.01^2$ $D_2 = \begin{bmatrix} 1^2 & 0 & \\ 0 & 1^2 & 0 \\ 0 & 0 & 1^2 \end{bmatrix}$
MHE 3	$\hat{x}_3(0) = [0.0693\ 0.3560\ 546.4699]^T$	10	$Q_3 = \begin{bmatrix} 0.001^2 & 0 & \\ 0 & 0.001^2 & 0 \\ 0 & 0 & 0.1^2 \end{bmatrix}$ $R_3 = 0.01^2$ $D_3 = \begin{bmatrix} 1^2 & 0 & \\ 0 & 1^2 & 0 \\ 0 & 0 & 1^2 \end{bmatrix}$

3.5.2 Unconstrained Case Results

In this case, we simulate the process without consideration of the inequality constraints to verify the convergence efficiency. The simulation results are shown in Figure 3.3 - Figure 3.6. Figure 3.3 shows the trajectories of the estimated states given by the centralized MHE, the decentralized MHE, the price-driven CDMHE and the actual state. It is shown that the estimated states given by the decentralized MHE are far from the actual states. Figure 3.4 is used to give us a clearer view of the estimated states given by the price-driven CDMHE and the centralized MHE. It can seen that the trajectory of estimated states given by the price-driven CDMHE is almost overlapped with the centralized trajectory, which means the price-driven CDMHE reaches the centralized performance. Figure 3.5 shows the trajectory of the error norm of the three estimated algorithms. It is shown that the performance of the centralized MHE and the proposed prediction-driven CDMHE algorithm are much better than the decentralized MHE. The summation of error term given by the price-driven CDMHE is 14.0345 while the summation of error term given by the decentralized MHE is 100.0387 which is around 7 times of the error norm provided by the price-driven CDMHE. Figure 3.6 shows the number of iterations during the sampling time intervals. It can be seen that the iterations are all 2, which verifies the theory in Section 3.3.

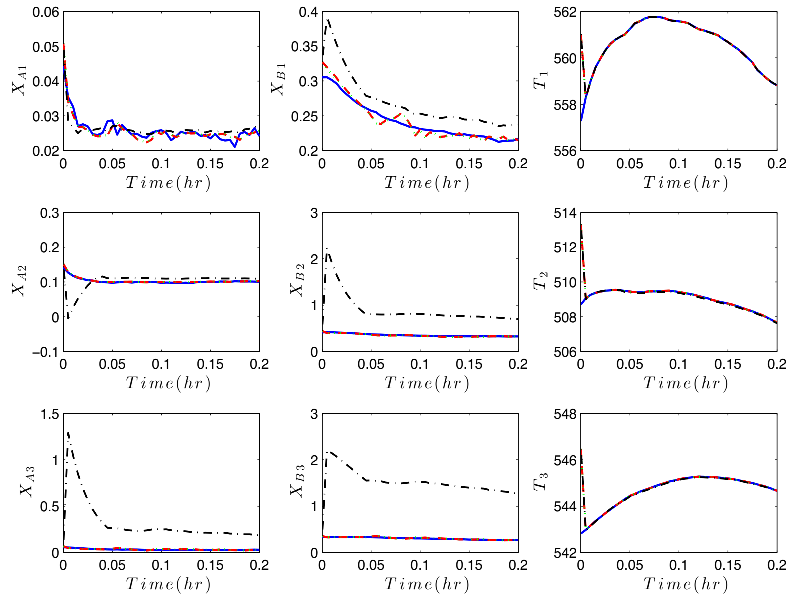

Figure 3.3: Trajectories of the actual states (solid line), estimates given by the proposed unconstrained price-driven CDMHE (dashed line), estimates given by the centralized MHE (dotted line), and the decentralized MHE (dash-dotted line).

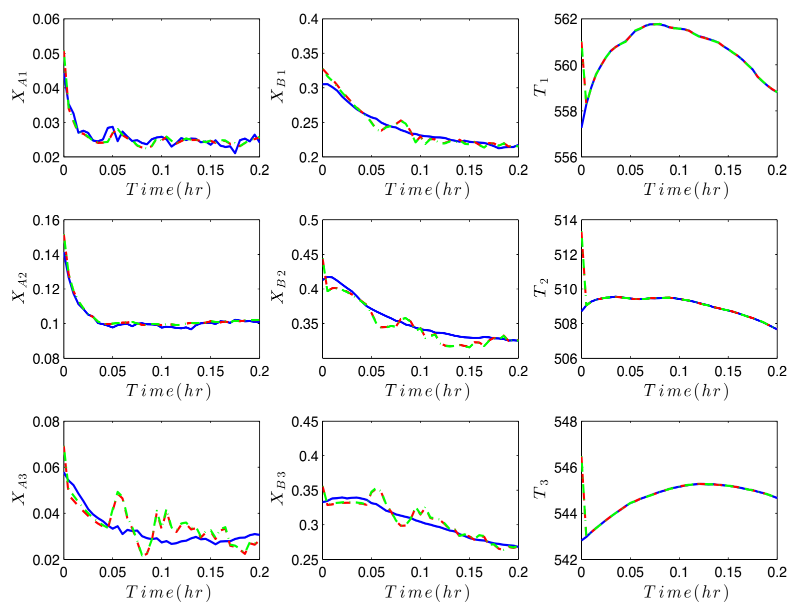

Figure 3.4: Trajectories of the actual states (solid line), estimates given by the proposed unconstrained price-driven CDMHE (dashed line), estimates given by the centralized MHE (dash-dotted line).

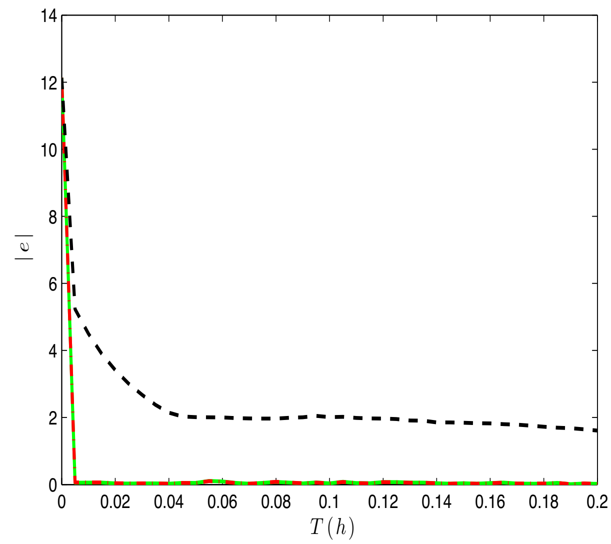

Figure 3.5: Trajectories of the estimation error norm given by the proposed unconstrained price-driven CDMHE (dash-dotted line), the centralized MHE (solid line), and the decentralized MHE (dashed line).

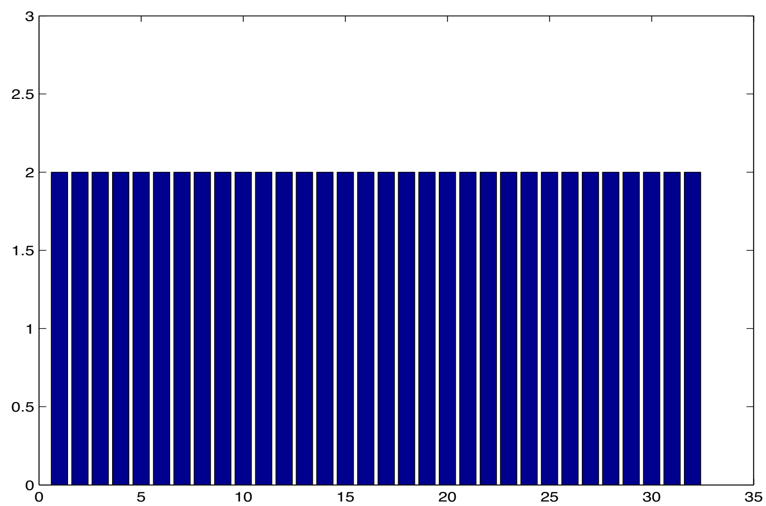

Figure 3.6: Numbers of iterations during each sampling time of the proposed unconstrained price-driven CDMHE.

3.5.3 Constrained Case Results

The constraints are taken into account by using the method described in Section 3.4. The simulation results are shown in Figure 3.7 - Figure 3.10. Figure 3.7 shows the trajectories of the estimated states given by the price-driven CDMHE, the centralized MHE, the decentralized MHE and the actual state. Because of the existence of constraints, the estimates given by the MHE formulations are different from the unconstrained case. Figure 3.8 is used to show a clearer view of the estimated states given by the constrained price-driven CDMHE and the constrained centralized MHE. It can seen that constrained price-driven CDMHE reaches the constrained centralized performance. Figure 3.9 shows the trajectory of the error norm of the three estimated algorithms. It is shown that the performance of the constrained centralized MHE and proposed constrained price-driven CDMHE algorithm are much better than constrained decentralized MHE. The summation of error term given by the constrained price-driven CDMHE is 13.9983 while the summation of error term given by the decentralized MHE is 142.0082 which around 10 times of the error norm provided by constrained price-driven CDMHE. Figure 3.6 shows the number of iterations during the sampling time intervals. It can be seen that the iterations are from 4 to 14 which verifies that the price-driven CDMHE algorithm with inequality constraints needs more iterations to converge.

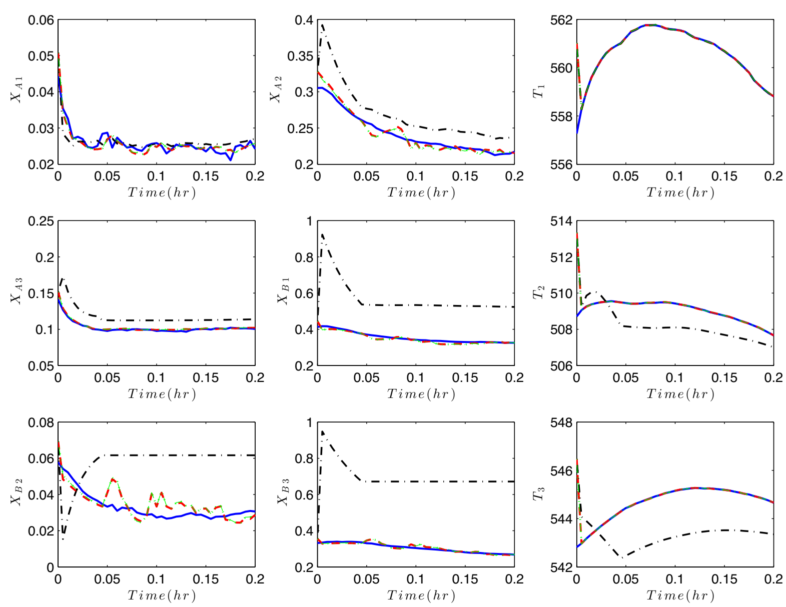

Figure 3.7: Trajectories of the actual states(solid line), estimates given by the proposed price-driven CDMHE (dashed line), estimates given by the centralized MHE (dotted line), and the decentralized MHE (dash-dotted line), all with inequality constraints.

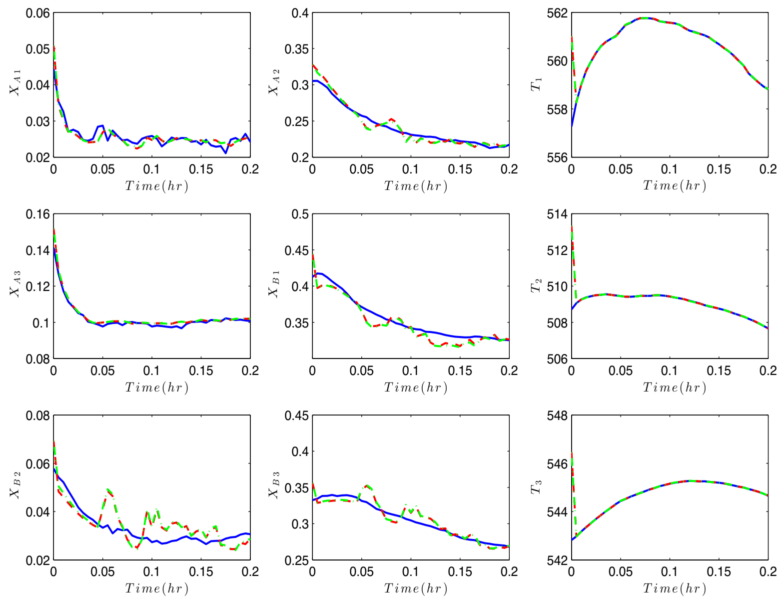

Figure 3.8: Trajectories of the actual states(solid line), estimates given by the proposed price-driven CDMHE (dashed line), estimates given by the centralized MHE (dash-dotted line) all with inequality constraints.

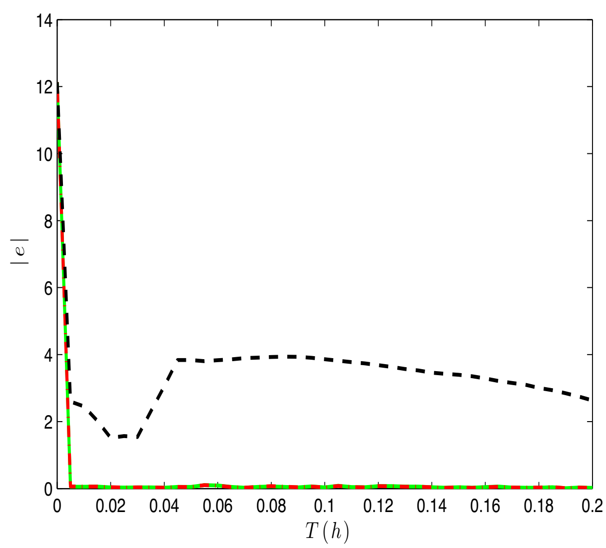

Figure 3.9: Trajectories of the estimation error norm given by the proposed price-driven CDMHE (dash-dotted line), the centralized MHE (solid line), and the decentralized MHE (dashed line) all with inequality constraints.

3.6 Conclusions

In this chapter, a price-driven coordinated algorithm is derived for the distributed moving horizon estimation, in which a local MHE estimates the process states, noises

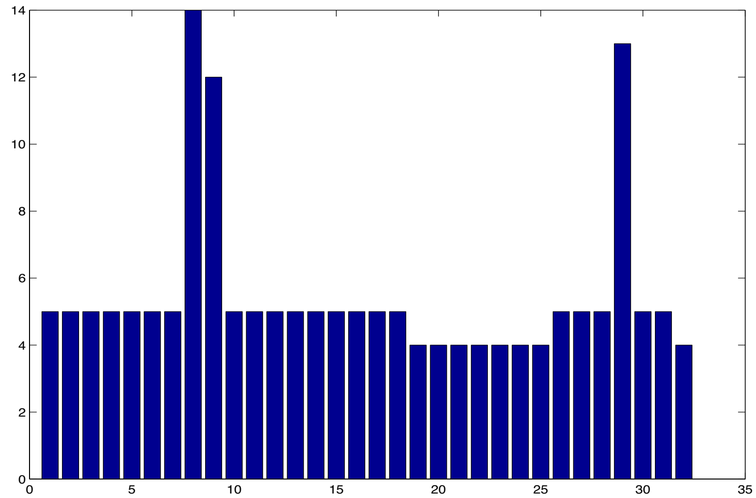

Figure 3.10: Number of iterations during each sampling time of the proposed constrained price-driven CDMHE

and interactions. It is shown that the conventional price-driven coordinated distributed moving horizon estimation(CDMHE) cannot be used for state estimation purpose since it requires measurement of the full state vector to ensure the existence of unique solution in the local MHEs. An improved price-driven CDMHE is firstly proposed to address the issue of the conventional version. An improved price-driven CDMHE is proposed without considering inequality constraints first. The analysis shows the unconstrained price-driven CDMHE converges to the performance of centralized MHE in two iterations during each sampling time. The simulation of a chemical process without inequality constraints verified this theory. The formulation of the improved price-driven CDMHE with inequality constraints is also described by dividing the inequality constraints into active constraints and inactive constraints. The simulation of the chemical process with inequality constraints shows good convergence to the centralized MHE.

The main strengths of the proposed improved price-driven CDMHE are:

- It is easier to solve than a global approach since only lower order subproblems are solved.
- The performance of the price-driven CDMHE reaches the corresponding centralized MHE.

- When Newton's method is used to update the price vector in the coordinator, the subsystems show fast convergence speed. Specially, without inequality constraints, the algorithm converges in two iterations.

However, the method also has certain drawbacks. The main disadvantage is the existence of the $\frac{1}{2} \parallel \hat{H}_i(k) \parallel_{\boldsymbol{D}_i^{-1}}$ in the cost function. This term is added purely to ensure unique solutions can be obtained by local MHEs. Nevertheless, by properly choosing the weighting matrix $\boldsymbol{D}_i^{-1}$, the effect of the additional term can be reduced. The proposed improved price-driven CDMHE is still an efficient distributed state estimation method.

Chapter 4

Prediction-driven Coordinated Distributed MHE

In this chapter, a prediction-driven coordinated moving horizon estimation algorithm is developed. The differences between the prediction-driven CDMHE and price-driven CDMHE exist in the coordinator and the estimated variables. The formulation of the unconstrained prediction-driven CDMHE is introduced first; then sufficient convergence conditions that ensure its convergence are provided; and a method to handle constraints is given at last.

4.1 Prediction-driven CDMHE Formulation

In this section, the prediction-driven CDMHE is presented. Figure 4.1 shows a schematic of the proposed design. In the proposed design, the local MHEs send subsystem state estimates to the coordinator; the coordinator calculates a price vector and the estimation of the interactions between subsystems and sends them to local MHEs. The above steps are carried out iteratively each sampling time. First, we introduce the subsystem MHE design. Subsequently, we present the coordinator design. Finally, we summarize the prediction-driven coordinated distributed state estimation algorithm.

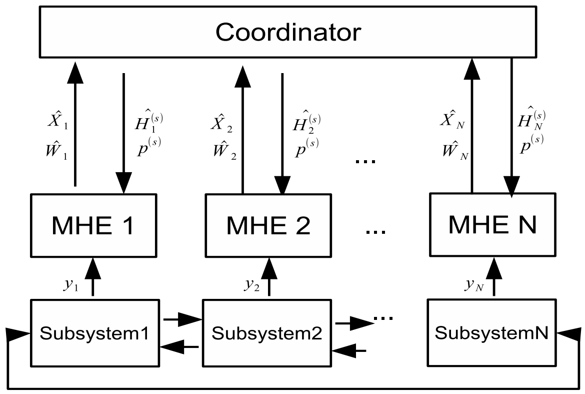

Figure 4.1: Information transfer in prediction-driven CDMHE with N subsystems

4.1.1 Subsystem MHE Design

In the coordinated distributed system, the model of the local subsystem i is as following:

$$\begin{aligned}
&\hat{\boldsymbol{x}}_i(k+1) = A_{ii}\hat{\boldsymbol{x}}_i(k) + \hat{\boldsymbol{h}}_i(k) + \hat{\boldsymbol{w}}_i(k);\\
&\hat{\boldsymbol{y}}_i(k) = C_{ii}\hat{\boldsymbol{x}}_i(k) + \hat{\boldsymbol{v}}_i(k)
\end{aligned} \tag{4.1}$$

where $\hat{\boldsymbol{h}}_i(k)$ denotes the estimated interaction of subsystem i with other subsystems which is calculated by the coordinator.

In this coordinated algorithm, a coordinating term is incorporated in the cost function of each local MHE. The coordinating term of MHE i ($i = 1, \ \ldots, \ N$) characterizes the effects of other subsystems on subsystem i. The coordinating term is also a link between the local estimator and the coordinator. For MHE i, the coordinating term is defined as $p^{(s)^T}\Theta_i \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \end{bmatrix}$ where $\hat{X}_i(k) = [\hat{\boldsymbol{x}}_i(k-Hp+1)^T, \ \ldots, \ \hat{\boldsymbol{x}}_i(k)^T]^T$, $\hat{W}_i(k) = [\hat{\boldsymbol{w}}_i(k-Hp+1)^T, \ \ldots, \ \hat{\boldsymbol{w}}_i(k-1)^T]^T$, and (4.3), (4.9) define Θ_i and $p^{(s)}$, respectively.

The local MHE for subsystem i at the s^{th} iteration is formulated as the following optimization problem:

$$\begin{aligned}
\min_{\hat{X}_i(k),\hat{W}_i(k)} \mathcal{J}_i =& \frac{1}{2}\begin{bmatrix}\hat{X}_i(k)^T & \hat{W}_i(k)^T\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & \\ & \boldsymbol{Q}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix}\\
&+\{\begin{bmatrix}-Y_i^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & 0\end{bmatrix} + \boldsymbol{p}^{(s)^T}\Theta_i\}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} \qquad (4.2a)\\
=&\frac{1}{2}Z_i(k)^T\Upsilon_i Z_i(k) + \{\Phi_i^T + \boldsymbol{p}^{(s)^T}\Theta_i\}Z_i(k)
\end{aligned}$$

$$\text{s.t.} \qquad G_i^{eq}Z_i(k) = \hat{H}_i^{(s)}(k) \tag{4.2b}$$

where $Y_i = [\boldsymbol{y}_i(k-Hp+1)^T, ..., \boldsymbol{y}_i(k)^T]^T$, $\hat{H}_i^{(s)}(k)$ is a variable determined by the coordinator which approximates the interaction between subsystem i and other subsystems, $\boldsymbol{p}^{(s)}$ is a vector determined by the coordinator and connects the local subsystem with the coordinator. The superscript 's' of $\hat{H}_i^{(s)}(k)$ and $\boldsymbol{p}^{(s)}$ indicates that they are coordinating variables calculated by the coordinator at the s^{th} iteration and communicated back to the local estimator. $G_i^{eq} = [G_{A_{ii}}, G_{B_{ii}}]$ with $G_{A_{ii}}$ and $G_{B_{ii}}$ are defined in (2.14) and (2.16). This coordinated design can guarantee the unique solution exists in the subsystem, the proof can be found in Appendix A.3.

In (4.2), Θ_i is a $(H_p-1)n \times (2H_p-1)n_i$ matrix built in the following block-wise fashion:

$$\Theta_i = \begin{bmatrix} G_{1i} \\ \vdots \\ 0 \\ \vdots \\ G_{Ni} \end{bmatrix} \begin{matrix} \\ i^{th} \text{ block is a} \\ \leftarrow \; (H_p-1)n_i \times (2H_p-1)n_i \\ \text{zero matrix.} \\ \end{matrix} \tag{4.3}$$

$\Theta_i = [\Theta_{A_i}, \Theta_{B_i}]$ with

$$\Theta_{A_i} = \begin{bmatrix} G_{A_{1i}} \\ \vdots \\ 0 \\ \vdots \\ G_{A_{Ni}} \end{bmatrix} \quad \Theta_{B_i} = \begin{bmatrix} G_{B_{1i}} \\ \vdots \\ 0 \\ \vdots \\ G_{B_{Ni}} \end{bmatrix} \tag{4.4}$$

4.1.2 Formulation of the Coordinator

Vector $\boldsymbol{p}^{(s)}$ in the local MHE design is a $(H_p-1)n \times 1$ vector which is an approximation of the centralized MHE Lagrange multiplier, computed and updated by the coordinator and will be referred to as the '*price vector*'. The price vector $p^{(s)}$ is same for all subsystems, therefore, it does not have subscript i.

At the s^{th} iteration of the sampling time k, the coordinator first sends $p^{(s)}$ and $\hat{H}_i^{(s)}(k)$ to local MHE i. Based on $p^{(s)}$ and $\hat{H}_i^{(s)}(k)$, the i^{th} local MHE problem (4.2) is solved and optimal estimates for the local state and disturbance $\hat{X}^*_{i(s)}(k)$ and $\hat{W}^*_{i(s)}(k)$ are obtained. The local MHE sends $\hat{X}^*_{i(s)}(k)$ and $\hat{W}^*_{i(s)}(k)$ to the coordinator. The 's' in the subscripts is used to denote that these optimal values are calculated in the s^{th} iteration by the local estimators.

The coordinator increases the iteration counter by one(i.e., $s \leftarrow s+1$), collects $\hat{X}^*_{i(s-1)}(k),\ i=1,\ \ldots,\ N$ and $\hat{W}^*_{i(s-1)}(k),\ i=1,...,N$, and assigns

$$\hat{X}_i^{(s)}(k) = \hat{X}^*_{i(s-1)}(k);\ \hat{X}^{(s)}(k) = \left[\hat{X}_1^{(s)}(k)^T, \cdots, \hat{X}_N^{(s)}(k)^T\right]^T. \tag{4.5a}$$

$$\hat{W}_i^{(s)}(k) = \hat{W}^*_{i(s-1)}(k);\ \hat{W}^{(s)}(k) = \left[\hat{W}_1^{(s)}(k)^T, \cdots, \hat{W}_N^{(s)}(k)^T\right]^T. \tag{4.5b}$$

The coordinator uses $\hat{X}^{(s)}(k)$ and $\hat{W}^{(s)}(k)$ to update estimated interaction $\hat{H}_i^{(s)}(k)$ and price vector $\boldsymbol{p}^{(s)}$. $\hat{H}_i^{(s)}(k)$ in equation (4.2) is the estimated interaction between subsystem i and other subsystems. It is defined as $\hat{H}_i^{(s)}(k) = [\hat{\boldsymbol{h}}_i^{(s)}(k - Hp + 1)^T,\ \ldots,\ \hat{\boldsymbol{h}}_i^{(s)}(k)^T]^T$. In this prediction-driven CDMHE algorithm, $\hat{H}_i(k)$ is calculated by the coordinator in the following equation:

$$\hat{H}_i^{(s)}(k) = -\sum_{j\neq i} G_{ij} \begin{bmatrix} \hat{X}_j^{(s)}(k) \\ \hat{W}_j^{(s)}(k) \end{bmatrix} \tag{4.6}$$

where $G_{ij} = \left[G_{A_{ij}}, G_{B_{ij}}\right]$, and $G_{A_{ij}}$, $G_{B_{ij}}$ are defined in (2.15) and (2.17).

The price vector is the approximation of the Lagrange multiplier in the centralized MHE, it can be obtained from following equation:

$$G^T\boldsymbol{p}^{(s)} = -\begin{bmatrix} \mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C} & \\ & \boldsymbol{Q}^{-1} \end{bmatrix} \begin{bmatrix} \hat{X}^{(s)}(k) \\ \hat{W}^{(s)}(k) \end{bmatrix} + \begin{bmatrix} \mathbb{C}^T\boldsymbol{R}^{-1}Y \\ 0 \end{bmatrix} \tag{4.7}$$

where $G = [G_A, G_B]$. (4.7) can be rewritten in the following form:

$$G_A^T\boldsymbol{p}^{(s)} = -\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C}\hat{X}^{(s)}(k) + \mathbb{C}^T\boldsymbol{R}^{-1}Y \tag{4.8a}$$

$$G_B^T\boldsymbol{p}^{(s)} = -\boldsymbol{Q}^{-1}\hat{W}^{(s)}(k) \tag{4.8b}$$

Since the size of matrix G is $(H_p-1)n \times (2H_p-1)n$, we can only get $\boldsymbol{p}^{(s)}$ by approximation. In this work, $\hat{X}^{(s)}(k)$ and $\hat{W}^{(s)}(k)$ are related through the equality constraint (4.2b). Since G_B is invertible, we pick equation (4.8b) to approximate $\boldsymbol{p}^{(s)}$, thus $\boldsymbol{p}^{(s)}$ can be calculated as:

$$\boldsymbol{p}^{(s)} = -{G_B^T}^{-1}\boldsymbol{Q}^{-1}\hat{W}^{(s)}(k) \tag{4.9}$$

$\hat{X}^{(s)}(k)$ is updated according to equation (4.5a). $\hat{W}^{(s)}(k)$ is updated according to equation (4.5b) and $\boldsymbol{p}^{(s)}$ is updated according to (4.9). After updating these variables, coordinator calculates interaction term $\hat{H}_i^{(s)}(k)$ for each subsystem i for

$i = 1,\ 2,\ \ldots,\ N$ according to (4.6). After that, coordinator sends $\boldsymbol{p}^{(s)}$ and $\hat{H}_i^{(s)}(k)$ to local estimators. The i^{th} local estimator receives $\boldsymbol{p}^{(s)}$ and $\hat{H}_i^{(s)}(k)$, and then solves the local optimization problem (4.2) to obtain the estimated state $\hat{X}_i^*$ and $\hat{W}_i^*$.

The above iteration process is terminated when the coordinator determines that $\| \hat{X}^{(s)}(k) - \hat{X}^{(s-1)}(k) \| < \epsilon$, where ϵ is pre-defined accuracy threshold.

The proposed prediction-driven CDMHE algorithm is summarized in Algorithm 2.

Algorithm 2 Implementation of Prediction-driven Coordinated Distributed MHE

Initialization

Coordinator: Set iteration counter $s = 0$. When $k = 1$, $\hat{X}^{(0)}(k)$ and $\hat{W}^{(0)}(k)$ are arbitrarily determined; when $k > 1$, $\hat{X}^{(0)}(k)$ and $\hat{W}^{(0)}(k)$ are picked as $\hat{X}(k-1)$, $\hat{W}(k-1)$. $\boldsymbol{p}^{(0)}(k)$ is calculated according to (4.9), $\hat{H}_i^{(0)}(k)$ is determined according to (4.6).

repeat

Coordinator: $\boldsymbol{p}^{(s)}$ and $\hat{H}_i^{(s)}(k)$ are sent to local MHE estimators.

Local Estimators: Problem (4.2) is solved. $\hat{X}_{i(s)}^*(k)$ and $\hat{W}_{i(s)}^*(k)$ are sent to the coordinator.

Coordinator: $s \leftarrow s + 1$, $\hat{X}^{(s)}(k)$ and $\hat{W}^{(s)}(k)$ are updated according to (4.5a) and (4.5b), $\boldsymbol{p}^{(s)}$ and $\hat{H}_i^{(s)}(k)$ are calculated based on (4.9) and (4.6).

until stopping criterion $\|\hat{X}^{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \epsilon$ is satisfied.

4.2 Convergence Performance Analysis

In this section, we first show that if at each sampling time the coordinated algorithm converges, the prediction-driven coordinated algorithm converges to the corresponding centralized performance. Then based on the iterative nature of the prediction-driven CDMHE, we derive a set of sufficient conditions under which the convergence of the algorithm is ensured.

4.2.1 Performance of the Coordinated Algorithm

In this section, we show that if the prediction-driven coordinated algorithm 2 converges every sampling time, the solution of the prediction-driven CDMHE converges to the centralized MHE.

Proposition 4.2.1. *Consider that the entire system described by (2.1) is estimated by N local coordinated DMHE estimators as given by solving (4.2). Suppose that the coordinated distributed MHE estimators are coordinated using the propose prediction-driven CDMHE algorithm 2 described in Section 4.1. When the CDMHE converges every sampling time, that is the stop criteria $\|\hat{X}^{(s)} - \hat{X}^{(s-1)}\| < \epsilon$ is satisfied, the solution obtained from CDMHE equals to the centralized optimal solutions.*

Proof Proposition 4.2.1 can be proven by comparing the centralized optimal solution to the solutions obtained with the prediction-driven CDMHE estimators.

Centralized Estimation Problem

As shown in (2.6), the centralized problem without inequality constraints can be written in terms of variable $Z(k)$ as follows:

$$\min_{Z(k)} \mathcal{J} = \frac{1}{2} Z(k)^T \Upsilon Z(k) + \Phi^T Z(k) \tag{4.10a}$$

subject to:

$$G^{eq} Z(k) = 0 \tag{4.10b}$$

where Υ and Φ are the same as defined in (2.6), $Z(k) = [\hat{X}(k)^T,\ \hat{W}(k)^T]^T$, $\hat{X}(k) = [\hat{X}_1(k)^T,\ \ldots,\ \hat{X}_N(k)^T]^T$ and $\hat{W}(k) = [\hat{W}_1(k)^T,\ \ldots,\ \hat{W}_N(k)^T]^T$, $G^{eq} = [G_A, G_B]$, G_A and G_B are defined in (2.13).

The Lagrange function of (4.10) is expressed as:

$$\mathcal{L}_{cen} = \frac{1}{2} Z(k)^T \Upsilon Z(k) + \Phi^T Z(k) + \lambda^T (G^{eq} Z(k)) \tag{4.11}$$

The optimal condition for a regular point $Z^*(k)$ to be a minimum in the quadratic problem (4.10) are:

$$\Upsilon Z^*(k) + \Phi + G^{eqT} \lambda^* = 0 \tag{4.12a}$$

$$G^{eq} Z^*(k) = 0 \tag{4.12b}$$

where $Z^*(k)$ is the optimal solution, λ^* is the Lagrange multiplier.

Coordinated Estimation Problem

Aggregating the subsystem optimization problem (4.2) from 1 to N, the aggregated form is expressed as follows:

$$\min_{\hat{X}(k),\hat{W}(k)} \mathcal{J} = \frac{1}{2}\begin{bmatrix}\hat{X}(k)^T & \hat{W}(k)^T\end{bmatrix}\begin{bmatrix}\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C} & 0\\ 0 & \boldsymbol{Q}^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}(k)\\ \hat{W}(k)\end{bmatrix} + \{\begin{bmatrix}-Y^T\boldsymbol{R}^{-1}\mathbb{C} & 0\end{bmatrix} + \boldsymbol{p}^{(s)^T}\bar{\Theta}\}\begin{bmatrix}\hat{X}(k)\\ \hat{W}(k)\end{bmatrix}$$
$$= \frac{1}{2}Z(k)^T\Upsilon Z(k) + \{\Phi^T + \boldsymbol{p}^{(s)^T}\bar{\Theta}\}Z(k) \tag{4.13a}$$

$$\text{subject to:} \quad \bar{G}^{eq}Z(k) = \hat{H}^{(s)}(k) \tag{4.13b}$$

where $\hat{X}(k)$ and $\hat{W}(k)$ are in the same order as in the centralized problem, Y, $\mathbb{C}$ and Φ are also same as in the centralized problem. $\bar{G}^{eq}$ is the aggregated form of G_i^{eq} and $\bar{G}^{eq} = [\bar{G}_A, \bar{G}_B]$. $\bar{\Theta}$ is the aggregated form of Θ_i and $\bar{\Theta} = [\bar{\Theta}_A, \bar{\Theta}_B]$. $\hat{H}^{(s)}(k)$ is the aggregated form of $\hat{H}_i^{(s)}(k)$ and $\hat{H}^{(s)}(k) = [\hat{H}_1^{(s)}(k)^T,\ \hat{H}_2^{(s)}(k)^T,\ \cdots,\ \hat{H}_N^{(s)}(k)^T]^T$. $\bar{G}_A, \bar{G}_B, \bar{\Theta}_A, \bar{\Theta}_B$ are respectively defined as:

$$\bar{G}_A = diag(G_{A_{11}}, ..., G_{A_{NN}}), \tag{4.14}$$

$$\bar{G}_B = diag(G_{B_{11}}, ..., \boldsymbol{G}_{B_{NN}}), \tag{4.15}$$

$$\bar{\Theta}_A = G_A - \bar{G}_A, \tag{4.16}$$

$$\bar{\Theta}_B = G_B - \bar{G}_B, \tag{4.17}$$

So we have:

$$\bar{\Theta} + \bar{G} = G \tag{4.18}$$

$$\hat{H}^{(s)}(k) = -\bar{\Theta}Z^{(s)}(k) \tag{4.19}$$

The first order optimal condition for equation (4.13) can be written as:

$$\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C}\hat{X}_{MHE}(k) - \mathbb{C}^T\boldsymbol{R}^{-1}\boldsymbol{Y} + \bar{\Theta}_A^T\boldsymbol{p}^{(s)} + \bar{G}_A^T\lambda_{MHE} = \emptyset \tag{4.20a}$$

$$\boldsymbol{Q}^{-1}\hat{W}_{MHE}(k) + \bar{\Theta}_B^T\boldsymbol{p}^{(s)} + \bar{G}_B^T\lambda_{MHE} = \emptyset \tag{4.20b}$$

$$\bar{G}_A\hat{X}_{MHE}(k) + \bar{G}_B\hat{W}_{MHE}(k) = -(\bar{\Theta}_A\hat{X}^{(s)}(k) + \bar{\Theta}_B\hat{W}^{(s)}(k)) \tag{4.20c}$$

where $\hat{X}_{MHE}(k) = [\hat{X}_1(k)^T,\ \cdots, \hat{X}_N(k)^T]^T$, $\hat{W}_{MHE}(k) = [\hat{W}_1(k)^T,\ \cdots, \hat{W}_N^T]^T$ and $\lambda_{MHE} = [\lambda_1^T,\ \cdots, \lambda_N^T]^T$. The subscript 'MHE' means that the solutions are the aggregation of the local MHE solutions.

The equation (4.20c) can be rewritten as following:

$$\bar{G}_A^{eq}\hat{X}_{MHE}(k) + \bar{G}_B^{eq}\hat{W}_{MHE}(k) + \bar{\Theta}_A\hat{X}^{(s)}(k) + \bar{\Theta}_B\hat{W}^{(s)}(k) = 0 \tag{4.21}$$

In our proposed coordinated algorithm, when the algorithm converges, $\hat{X}^{(s)}(k)$ almost equals to $\hat{X}^{(s-1)}(k)$. This implies that in the last iteration, $\hat{X}^{(s)}(k) = \hat{X}_{MHE}(k)$ and $\hat{W}^{(s)}(k) = \hat{W}_{MHE}(k)$. Based on this, equation (4.21) can be written as:

$$\bar{G}_A^{eq}\hat{X}_{MHE}(k) + \bar{G}_B^{eq}\hat{W}_{MHE}(k) + \bar{\Theta}_A\hat{X}_{MHE}(k) + \bar{\Theta}_B\hat{W}_{MHE}(k) = 0 \tag{4.22}$$

So the first order optimal conditions of prediction-driven CDMHE (4.20) becomes:

$$\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C}\hat{X}_{MHE}(k) - \mathbb{C}^T\boldsymbol{R}^{-1}Y + \bar{\Theta}_A^T\boldsymbol{p}^{(s)} + \bar{G}_A^T\lambda_{MHE} = 0 \tag{4.23a}$$

$$\boldsymbol{Q}^{-1}\hat{W}_{MHE}(k) + \bar{\Theta}_B^T\boldsymbol{p}^{(s)} + \bar{G}_B^T\lambda_{MHE} = 0 \tag{4.23b}$$

$$\bar{G}_A\hat{X}_{MHE}(k) + \bar{G}_B\hat{W}_{MHE}(k) + \bar{\Theta}_A\hat{X}_{MHE}(k) + \bar{\Theta}_B\hat{W}_{MHE}(k) = 0 \tag{4.23c}$$

Given $G_{B_{ij}} = 0$ is defined in (2.17), G_B and $\bar{G}_B$ as defined in (2.13) and (4.15) respectively, the following equation can be concluded:

$$G_B = \bar{G}_B \tag{4.24}$$

Given the definition of $\bar{\Theta}_B$ as in equation (4.17), we can get $\bar{\Theta}_B = 0$. The equation (4.23b) can be written as:

$$G_B^T\lambda_{MHE} = -\boldsymbol{Q}^{-1}\hat{W}_{MHE}(k) \tag{4.25}$$

The price vector $\boldsymbol{p}^{(s)}$ is updated by:

$$\boldsymbol{p}^{(s)} = -{G_B^T}^{-1}\boldsymbol{Q}^{-1}\hat{W}^{(s)}(k) \tag{4.26}$$

Since at the end of iteration, $\hat{W}^{(s)}(k) = \hat{W}_{MHE}(k)$, comparing (4.25) and (4.26), we can get that $\boldsymbol{p}^{(s)} = \lambda_{MHE}$. Therefore, at the end of iteration, equation (4.23) can be rewritten as:

$$\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C}\hat{X}_{MHE}(k) - \mathbb{C}^T\boldsymbol{R}^{-1}Y + \bar{\Theta}_A^T\lambda_{MHE} + \bar{G}_A^T\lambda_{MHE} = 0 \tag{4.27a}$$

$$\boldsymbol{Q}^{-1}\hat{W}_{MHE}(k) + \bar{G}_B^T\lambda_{MHE} = 0 \tag{4.27b}$$

$$\bar{G}_A\hat{X}_{MHE}(k) + \bar{G}_B\hat{W}_{MHE}(k) + \bar{\Theta}_A\hat{X}_{MHE}(k) + \bar{\Theta}_B\hat{W}_{MHE}(k) = 0 \tag{4.27c}$$

Let us denote $Z_{MHE}(k) = [\hat{X}_{MHE}^T,\ \hat{W}_{MHE}^T]^T$, and write the equation (4.27) into terms of Z_{MHE}. The following equation can be obtained:

$$\Upsilon Z_{MHE}(k) + \Phi + G^{eqT}\lambda_{MHE} = 0 \tag{4.28a}$$

$$G^{eq}Z_{MHE}(k) = 0 \tag{4.28b}$$

Comparing the first order optimal conditions of the centralized problem (4.12) with the ones of the coordinated problem (4.28) at the end of iteration, we can see that they are all same except the name of variables. The solution of (4.12) and (4.28) must be same; this means that at the end of the iteration, the estimated states of the proposed prediction-driven CDMHE $Z_{MHE}(k)$ converges to the centralized estimate $Z^*(k)$. □

4.2.2 Convergence Conditions

In this section, we provide sufficient conditions to ensure convergence of the proposed prediction-driven coordinated algorithm. The sufficient conditions given in this section ensure the convergence of the proposed prediction-driven CDMHE algorithm to the centralized MHE. The proof of the convergence is inspired by the convergence analysis of coordinated continuous-time linear quadratic regulators in Cohen, 1977.

First, we define the aggregated cost function without coordinating terms of the CDMHE J_D as follows:

$$\mathcal{J}_D = \sum_{i=1}^{N} \mathcal{J}_{Di} \tag{4.29}$$

where

$$\begin{aligned} \mathcal{J}_{Di} &= \frac{1}{2}\left(|Y_i(k) - C_{ii}\hat{X}_i(k)|^2_{R_i^{-1}} + |\hat{W}_i(k)|^2_{Q_i^{-1}}\right) \\ s.t.\ & G_{A_{ii}}\hat{X}_i(k) + G_{B_{ii}}\hat{W}_i(k) = \hat{H}_i(k) \\ \hat{H}_i(k) &= -\sum_{j\neq i} G_{ij}\begin{bmatrix}\hat{X}_j(k)\\ \hat{W}_j(k)\end{bmatrix} \end{aligned} \tag{4.30}$$

where $\mathcal{J}_{Di}$ is the cost function in local MHE i of (4.2) without the coordinating term, $\hat{H}_i(k)$ is the accurate interaction in subsystem i.

Based on the subsystem model, we are able to express $\hat{W}_i$ at a specific time instant k in terms of the subsystem state and interaction as:

$$\begin{aligned} &\hat{W}_i(k) = G_{B_{ii}}^{-1}(\hat{H}_i(k) - G_{A_{ii}}\hat{X}_i(k)) = f_i(\hat{X}_i(k), \hat{H}_i(k)) \\ &for: \ i = 1,\ 2,\ \cdots,\ N. \end{aligned} \tag{4.31}$$

So the objective function $\mathcal{J}_{Di}$ can be written as follows:

$$\mathcal{J}_{Di}(\hat{X}_i(k), \hat{H}_i(k)) = \frac{1}{2}\left(|Y_i(k) - C_{ii}\hat{X}_i(k)|^2_{R_i^{-1}} + |f_i(\hat{X}_i(k), \hat{H}_i(k))|^2_{Q_i^{-1}}\right) \tag{4.32}$$

Therefore, the objective function $\mathcal{J}_D$ defined in (4.29) can be rewritten as:

$$\mathcal{J}_D(\hat{X}(k), \hat{H}(k)) = \frac{1}{2}\sum_{i=1}^{N}\left(|Y_i(k) - \boldsymbol{C}_{ii}\hat{X}_i(k)|^2_{R_i^{-1}} + |f_i(\hat{X}_i(k), \hat{H}_i(k))|^2_{Q_i^{-1}}\right) \tag{4.33}$$

Define $\hat{W}_{DMHE}$ as the vector that contains all the subsystem $\hat{W}_i(k)$, $\hat{W}_{DMHE}(k) = [\hat{W}_1(k)^T,\ \hat{W}_2(k)^T, ...,\ \hat{W}_N(k)^T]^T$. Then according to the relationship between $\hat{W}_i(k)$ and $\hat{X}_i(k),\ \hat{H}_i(k)$ in (4.31), $\hat{W}_{DMHE}(k)$ can be written as:

$$\hat{W}_{DMHE}(k) = \bar{G}_B^{-1}(\hat{H}(k) - \bar{G}_A\hat{X}(k)) = f(\hat{X}(k), \hat{H}(k)) \tag{4.34}$$

where $\hat{X}(k) = [\hat{X}_1(k)^T, \hat{X}_2(k)^T, ..., \hat{X}_N(k)^T]^T$ and $\hat{H}(k) = [\hat{H}_1(k)^T, \hat{H}_2(k)^T, ..., \hat{H}_N(k)^T]^T$, $\bar{G}_A$ and $\bar{G}_B$ are defined in (4.14) and (4.15) respectively.

Let K be the mapping from state $\hat{X}(k)$ to the interaction $\hat{H}(k)$. Then $\hat{W}_{DMHE}(k)$ can be rewritten as: $\hat{W}_{DMHE}(k) = f(\hat{X}(k), K(\hat{X}(k)))$. The objective equation $\mathcal{J}_D$ can be expressed as follows:

$$\mathcal{J}_D(\hat{X}(k), \hat{H}(k)) = \frac{1}{2}\left(|Y(k) - \mathbb{C}\hat{X}(k)|^2_{\boldsymbol{R}^{-1}} + |f(\hat{X}(k), K(\hat{X}(k)))|^2_{\boldsymbol{Q}^{-1}}\right) \tag{4.35}$$

For the centralized MHE (2.6), the estimated noise $\hat{W}(k)$ can also be expressed as a function of states:

$$\hat{W}(k) = -G_B^{-1}G_A\hat{X}(k) = \bar{f}(\hat{X}(k)) \tag{4.36}$$

where the vector $\hat{W}(k)$ is arranged in the same way as the vector $\hat{W}_{DMHE}(k)$. The relationship between $\hat{W}(k)$ and $\hat{X}(k)$ can be derived from (2.6b) which gives a linear relationship.

The centralized objective function (2.6) can be expressed in terms of the estimated states as:

$$\mathcal{J}_{cen}(\hat{X}(k)) = \frac{1}{2}\left(|Y(k) - \mathbb{C}\hat{X}(k)|^2_{\boldsymbol{R}^{-1}} + |\bar{f}(\hat{X}(k))|^2_{\boldsymbol{Q}^{-1}}\right) \tag{4.37}$$

The objective functions $\mathcal{J}_D(\hat{X}(k), \hat{H}(k))$ and $\mathcal{J}_{cen}(\hat{X}(k))$ are both quadratic functions that can be differentiated to obtain:

$$\Psi = \frac{d^2 J_{cen}}{dX(k)^2} \tag{4.38a}$$

$$\Omega = diag(\Omega_i) \tag{4.38b}$$

where $\Omega_i = \dfrac{\partial^2 \mathcal{J}_D}{\partial \hat{X}_i^2(k)}$, for $i = 1, 2, ..., N$. The matrices Ψ and Ω will be used in the following theorem in characterizing the convergence conditions.

Theorem 4.2.1. *Consider system (2.1) with the proposed prediction-driven coordinated distributed state estimation algorithm described in Section 4.1. If $\Lambda = (\Omega - \frac{\Psi}{2})$ is positive definite, then the estimates given by the prediction-driven CDMHE algorithm converge to the ones obtained by the centralized MHE of (2.6) at each sampling time as the iteration number s increases.*

Proof For any given estimated state $\hat{X}(k)$:

$$\bar{f}(\hat{X}(k)) = -G_B^{-1} G_A \hat{X}(k) \tag{4.39}$$

According to (4.30), we can get the aggregated form $\hat{H}(k)$ as following form:

$$\hat{H}(k) = -\begin{bmatrix}\bar{\Theta}_A & \bar{\Theta}_B\end{bmatrix}\begin{bmatrix}\hat{X}(k) \\ \hat{W}(k)\end{bmatrix} \tag{4.40}$$

while $\Theta_i = [\Theta_{iA},\ \Theta_{iB}]$, Θ_{iA} and Θ_{iB} are defined in (4.4), $\bar{\Theta}_A = [\Theta_{1A},\ \Theta_{2A}, \cdots,\ \Theta_{NA}]$, $\bar{\Theta}_B = [\Theta_{1B},\ \Theta_{2B}, \cdots,\ \Theta_{NB}]$. Since $G_{B_{ij}} = 0$ from (2.17), thus $\bar{\Theta}_B = 0$, $\bar{G}_B = G_B$. From (4.14) and the definition of $\bar{\Theta}_A$, it can be obtained that:

$$G_A = \bar{G}_A + \bar{\Theta}_A \tag{4.41}$$

Thus for any given estimated state $\hat{X}(k)$:

$$\begin{aligned} f(\hat{X}(k), \hat{H}(k)) &= \bar{G}_B^{-1}(\hat{H}(k) - \bar{G}_A \hat{X}(k)) \\ &= \bar{G}_B^{-1}(-\bar{\Theta}_A \hat{X}(k) - \bar{G}_A \hat{X}(k)) \\ &= -G_B^{-1} G_A \hat{X}(k) \end{aligned} \tag{4.42}$$

So we get the following equation:

$$\bar{f}(\hat{X}(k)) = f(\hat{X}(k), K(\hat{X}(k))) \tag{4.43}$$

which also implies that:

$$\mathcal{J}_{cen}(\hat{X}(k)) = \mathcal{J}_D(\hat{X}(k), \hat{H}(k)) \tag{4.44}$$

Evaluating the first order derivative respect to state $\hat{X}(k)$ at both sides of the above equation (4.44), the following equation can be obtained:

$$\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)} = \frac{\partial \mathcal{J}_D}{\partial \hat{X}(k)} + \frac{\partial \mathcal{J}_D}{\partial \hat{H}(k)} \frac{d\hat{H}(k)}{d\hat{X}(k)} \tag{4.45}$$

It can be seen that the objective function of prediction-driven CDMHE (4.2) $\mathcal{J}_i$ can be rewritten as:

$$\mathcal{J}_i(\hat{X}_i(k)) = \mathcal{J}_{Di}(\hat{X}_i(k), \hat{H}_i^{(s)}(k)) + \boldsymbol{p}^{(s)T} \Theta_i \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \end{bmatrix} \tag{4.46}$$

For the subsystem optimization problem (4.2), it can be verified based on the definitions of the matrices in the cost function and system description that

$$\boldsymbol{p}^{(s)T} \Theta_i \begin{bmatrix} \hat{X}_i(k) \\ \hat{W}_i(k) \end{bmatrix} = \mathit{\Gamma}_i(\hat{X}^{(s)}(k))\hat{X}_i(k) \tag{4.47}$$

while

$$\mathit{\Gamma}_i(\hat{X}^{(s)}(k)) = \frac{\partial \mathcal{J}_D}{\partial \hat{H}(k)}|_{(\hat{X}^{(s)}(k), K(\hat{X}^{(s)}(k)))} \frac{d\hat{H}(k)}{d\hat{X}_i(k)}|_{\hat{X}^{(s)}(k)} \tag{4.48}$$

The proof of equation (4.47) requires somewhat lengthy calculations, which are given in Appendix B. The equation (4.47) and equation (4.45) reveal the relation between the centralized MHE and the proposed CDMHE design.

Now, Let us consider the actual cost function used in the CDMHE design and denote $\mathcal{J}_{CDMHE}$ as the aggregation of the subsystem cost function such that:

$$\mathcal{J}_{CDMHE} = \sum_{i=1}^{N} \mathcal{J}_i \tag{4.49}$$

It can be calculated that:

$$\frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)}|_{\hat{X}^{(s)}(k)} = \frac{\partial \mathcal{J}_D}{\partial \hat{X}(k)}|_{\hat{X}^{(s)}(k)} + \frac{\partial \mathcal{J}_D}{\partial \hat{H}(k)} \frac{d\hat{H}(k)}{d\hat{X}(k)}|_{\hat{X}^{(s)}(k)} \tag{4.50}$$

According to (4.45) and (4.50), $\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)} = \frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)}$ when they are evaluated at $\hat{X}^{(s)}(k)$. This result will be used in the next part of the proof.

Next, let us consider the Lagrangian functions of the centralized MHE and the CDMHE problems. Note that the equality constraint in (2.6) has already been used in the expression of $\hat{W}$ in terms of $\hat{X}$, so the Lagrangian function is equal to the objective function.

If we denote the optimal solution to the centralized MHE at time instant k as $\hat{X}^*(k)$ and perform Taylor series expansion of $\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)}$ around $\hat{X}^*(k)$, it can be obtained that:

$$\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)} = \frac{d\mathcal{J}_{cen}}{d\hat{X}(k)}|_{\hat{X}^*(k)} + \frac{d^2\mathcal{J}_{cen}}{d\hat{X}^2(k)}|_{\hat{X}^*(k)}(\hat{X}(k) - \hat{X}^*(k)) \tag{4.51}$$

Note that since the objective function $\mathcal{J}_{cen}$ is quadratic, so the higher terms in (4.51) are zeros. Also, since $\hat{X}^*(k)$ is the optimal solution of the centralized MHE at time k, we can get

$$\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)}|_{\hat{X}^*(k)} = 0 \tag{4.52}$$

Moreover, according to (4.38), (4.51) can be rewritten as:

$$\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)} = \Psi(\hat{X}(k) - \hat{X}^*(k)) \tag{4.53}$$

If we evaluate $\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)}$ at $\hat{X}^{(s)}(k)$, we can get:

$$\frac{d\mathcal{J}_{cen}}{d\hat{X}(k)}|_{\hat{X}^{(s)}(k)} = \Psi(\hat{X}^{(s)}(k) - \hat{X}^*(k)) \tag{4.54}$$

Let us also perform a Taylor series expansion of $\frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)}$ around $\hat{X}^{(s+1)}(k)$ which is the optimal solution of the CDMHE at the end of s^{th} iteration, we have:

$$\frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)} = \frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)}|_{\hat{X}^{(s+1)}(k)} + \frac{d^2\mathcal{J}_{CDMHE}}{d\hat{X}^2(k)}|_{\hat{X}^{(s+1)}(k)}(\hat{X}(k) - \hat{X}^{(s+1)}(k)) \tag{4.55}$$

Because $\hat{X}^{(s+1)}(k)$ is the optimal solution of the CDMHE at s^{th} iteration, we can get that

$$\frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)}|_{\hat{X}^{(s+1)}(k)} = 0 \tag{4.56}$$

Form equation (4.47), it can be observed that the coordinating term in $\mathcal{J}_i$ is a linear function of $\hat{X}_i(k)$, thus $\frac{d^2\mathcal{J}_{CDMHE}}{d\hat{X}^2(k)} = \frac{d^2\mathcal{J}_D}{d\hat{X}^2(k)} = diag(\frac{d^2\mathcal{J}_D}{d\hat{X}_i^2(k)}) = \Omega$. Equation (4.55) can be further rewritten as:

$$\frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)} = \Omega(\hat{X}(k) - \hat{X}^{(s+1)}(k)) \tag{4.57}$$

If we evaluate $\frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)}$ at $\hat{X}^{(s)}(k)$, we can get:

$$\frac{d\mathcal{J}_{CDMHE}}{d\hat{X}(k)}|_{\hat{X}^{(s)}(k)} = \Omega(\hat{X}^{(s)}(k) - \hat{X}^{(s+1)}(k)) \tag{4.58}$$

From (4.45), (4.50), (4.54) and (4.58), we get:

$$\hat{X}^{(s)}(k) - \hat{X}^{(s+1)}(k) = \Omega^{-1}\Psi(\hat{X}^{(s)}(k) - \hat{X}^{*}(k)). \tag{4.59}$$

Now, let us expand $\mathcal{J}_{cen}(\hat{X}(k))$ around $\hat{X}^{(s)}(k)$; that is,

$$\begin{aligned}\mathcal{J}_{cen}(\hat{X}(k)) =& \mathcal{J}_{cen}(\hat{X}^{(s)}(k)) + \frac{d\mathcal{J}_{cen}}{d\hat{X}(k)}|^T_{\hat{X}^{(s)}(k)}(\hat{X}(k) - \hat{X}^{(s)}(k)) \\ &+ (\hat{X}(k) - \hat{X}^{(s)}(k))^T \frac{1}{2}\frac{d^2\mathcal{J}_{cen}}{d\hat{X}^2(k)}|^T_{\hat{X}^{(s)}(k)}(\hat{X}(k) - \hat{X}^{(s)}(k))\end{aligned} \tag{4.60}$$

Evaluating (4.60) at $\hat{X}^{(s+1)}(k)$, we can get:

$$\begin{aligned}\mathcal{J}_{cen}(\hat{X}^{s+1}(k)) =& \mathcal{J}_{cen}(\hat{X}^{s}(k)) + (\hat{X}^{s}(k) - \hat{X}^{s+1}(k))^T\Omega(\hat{X}^{s+1}(k) - \hat{X}^{s}(k)) \\ &+ (\hat{X}^{s+1}(k) - \hat{X}^{s}(k))^T\frac{\Psi}{2}(\hat{X}^{s+1}(k) - \hat{X}^{s}(k))\end{aligned} \tag{4.61}$$

From equation (4.61) and (4.59), we can get the expression of $\mathcal{J}_{cen}(\hat{X}^{s}(k)) - \mathcal{J}_{cen}(\hat{X}^{s+1}(k))$ as following:

$$\mathcal{J}_{cen}(\hat{X}^{(s)}(k)) - \mathcal{J}_{cen}(\hat{X}^{(s+1)}(k)) = (\hat{X}^{(s)}(k) - \hat{X}^{*}(k))^T\Psi^T(\Omega^{-1})^T(\Omega - \frac{\Psi}{2})\Omega^{-1}\Psi(\hat{X}^{(s)}(k) - \hat{X}^{*}(k)) \tag{4.62}$$

We can make the following conclusions: If $\Omega - \frac{\Psi}{2} > 0$ is satisfied, we can get that $\mathcal{J}_{cen}(\hat{X}^{(s)}(k)) > \mathcal{J}_{cen}(\hat{X}^{(s+1)}(k))$. This implies that with the increasing of iteration s, the estimate obtained by the prediction-driven CDMHE $\hat{X}^{(s)}(k)$ converges to the centralized estimate $\hat{X}^{*}(k)$. □

4.3 Prediction-driven CDMHE Formulation with Inequality Constraints

One main advantage of moving horizon estimation is the ability to handle constraints. In Section 4.1, the prediction-driven coordinated distributed moving horizon estimation algorithm is formulated without inequality constraints. In this section, we propose to use barrier functions to handle the inequality constraints. In this way, we transfer the inequality constraints into the cost function. Specially, we consider the constraints on the estimated state $\hat{X}$ and the estimated process noise $\hat{W}$.

The objective function of the prediction-driven CDMHE becomes:

$$\min_{\hat{X}_i(k),\hat{W}_i(k)} \mathcal{J}_i = \frac{1}{2}\left[\hat{X}_i(k)^T \; \hat{W}_i(k)^T\right]\begin{bmatrix}\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & \\ & \boldsymbol{Q}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} + \{\left[-Y_i^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} \quad 0\right] + \boldsymbol{p}^{(s)^T}\Theta_i\}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} \tag{4.63a}$$

$$\text{s.t.} \quad G_{A_{ii}}\hat{X}_i(k) + G_{B_{ii}}\hat{W}_i(k) = \hat{H}_i^{(s)}(k), \qquad \hat{X}_i(k) \in \mathbb{X}, \quad \hat{W}_i(k) \in \mathbb{W} \tag{4.63b}$$

First, we introduce the barrier function, $\rho : \boldsymbol{R} \to \boldsymbol{R}$ as follows:

$$\rho(u) = \begin{cases} 0 & lb \le u \le ub \\ +\infty & else \end{cases} \tag{4.64}$$

where lb is the lower bound of u and ub is the upper bound of u. From the definition of the barrier function, it can be seen that the barrier function acts like a wall to constrain the variable into the boundary.

Our goal is to approximately formulate the inequality constrained prediction-driven CDMHE as an unconstrained problem. Using the barrier function introduced in equation (4.64), the augmented objective function of local MHE i in the prediction-driven CDMHE denoted as $\mathcal{J}_{ic}$ can be rewritten as:

$$\min_{\hat{X}_i(k),\hat{W}_i(k)} \mathcal{J}_{ic} = \frac{1}{2}\left[\hat{X}_i(k)^T\ \hat{W}_i(k)^T\right]\begin{bmatrix}\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & \\ & \boldsymbol{Q}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} + \{\left[-Y_i^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii}\quad 0\right] + {\boldsymbol{p}^{(s)}}^T\Theta_i\}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} + \rho_{ix}(\hat{X}_i(k)) + \rho_{iw}(\hat{W}_i(k)) \tag{4.65a}$$

$$\text{s.t.}\qquad G_{A_{ii}}\hat{X}_i(k) + G_{B_{ii}}\hat{W}_i(k) = \hat{H}_i^{(s)}(k) \tag{4.65b}$$

where

$$\rho_{ix}(\hat{X}_i(k)) = \sum_{l=K-H_p+1}^{K}\sum_{j=1}^{n_i}\rho(\hat{x}_{ij}(l)); \tag{4.66a}$$

$$\rho_{iw}(\hat{W}_i(k)) = \sum_{l=K-H_p+1}^{K-1}\sum_{j=1}^{n_i}\rho(\hat{w}_{ij}(l)) \tag{4.66b}$$

where $\hat{X}_i(k) = [\hat{x}_i(k-H_p+1)^T,\ \hat{x}_i(k-Hp+2)^T,\ \cdots,\ \hat{x}_i(k)^T]^T$ for $i = 1,\ 2,\ \cdots,\ N$. $\hat{x}_i(l) = [\hat{x}_{i1}(l)^T,\ \hat{x}_{i2}(l)^T,\ \cdots, \hat{x}_{in_i}(l)^T]^T$ and n_i is the number of states in subsystem i. Let us denote this kind prediction-driven CDMHE as augmented prediction-driven CDMHE.

Therefore the aggregated augmented objective function of constrained prediction-driven CDMHE denoted as $\mathcal{J}_{CDMHE}$ can be obtained as:

$$\mathcal{J}_{CDMHE} = \sum_{i=1}^{N}\mathcal{J}_{ic} \tag{4.67}$$

For the centralized MHE, the same method is applied to transfer the inequality constraints on estimated states $\hat{X}$ and estimated process noise $\hat{W}$ into the objective function. Let us denote this kind of centralized MHE as augmented centralized MHE. Thus the objective function of augmented centralized MHE has the following expression:

$$\min_{\hat{X}(k),\hat{W}(k)} \mathcal{J}_{cenc} = \frac{1}{2}\left[\hat{X}(k)^T\quad \hat{W}(k)^T\right]\begin{bmatrix}\mathbb{C}^T\boldsymbol{R}^{-1}\mathbb{C} & \\ & \boldsymbol{Q}^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}(k)\\ \hat{W}(k)\end{bmatrix} + \left[-Y^T\boldsymbol{R}^{-1}\mathbb{C}\quad 0\right]\begin{bmatrix}\hat{X}(k)\\ \hat{W}(k)\end{bmatrix} + \rho_x(\hat{X}(k)) + \rho_w(\hat{W}(k)) \tag{4.68a}$$

$$\text{s.t.}\quad G_A\hat{X}(k) + G_B\hat{W}(k) = 0 \tag{4.68b}$$

where

$$\rho_x(\hat{X}(k)) = \sum_{i=1}^{N}\{\sum_{l=K-H_p+1}^{K}\sum_{j=1}^{n_i}\rho(\hat{x}_{ij}(l))\} \tag{4.69a}$$

$$\rho_w(\hat{W}(k)) = \sum_{i=1}^{N}\{\sum_{l=K-H_p+1}^{K-1}\sum_{j=1}^{n_i}\rho(\hat{w}_{ij}(l))\} \tag{4.69b}$$

$$\tag{4.69c}$$

where $\hat{X}(k) = [\hat{X}_1(k)^T,\ \hat{X}_2(k)^T,\ \cdots, \hat{X}_N(k)^T]^T$, $\hat{X}_i(k)$ is arranged in the same way as the subsystem i in the prediction-driven CDMHE.

The same barrier function is used to transfer the inequality constraints into the cost function in the decentralized MHE. Let us denote this kind decentralized MHE as augmented decentralized MHE. For subsystem i described in (2.19), the new augmented objective function denoted as $\mathcal{J}_{idc}$ becomes:

$$\min_{\hat{X}_i(k),\hat{W}_i(k)} \mathcal{J}_{idc} = \frac{1}{2}\begin{bmatrix}\hat{X}_i(k)^T & \hat{W}_i(k)^T\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_{ii}^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & \\ & \boldsymbol{Q}_i^{-1}\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} + \begin{bmatrix}-Y_i^T\boldsymbol{R}_i^{-1}\boldsymbol{C}_{ii} & 0\end{bmatrix}\begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix}$$
$$+\rho_{ix}(\hat{X}_i(k)) + \rho_{iw}(\hat{W}_i(k)) \tag{4.70a}$$

$$\text{s.t.} \quad G_{A_{ii}}\hat{X}_i(k) + G_{B_{ii}}\hat{W}_i(k) = 0 \tag{4.70b}$$

where $\rho_{ix}(\hat{X}_i(k))$ and $\rho_{iw}(\hat{W}_i(k))$ are defined in equation (4.66a) and equation (4.66b) respectively.

The corresponding price vector in the augmented prediction-driven CDMHE is calculated as:

$$\boldsymbol{p}^{(s)} = -G_B^{T^{-1}}\boldsymbol{Q}^{-1}\hat{W}^{(s)}(k) - \frac{d\rho_w(\hat{W}^{(s)}(k))}{d\hat{W}^{(s)}(k)} \tag{4.71}$$

where

$$\frac{d\rho_w(\hat{W}^{(s)}(k))}{d\hat{W}^{(s)}(k)} = [\frac{d\rho_{1w}(\hat{W}_1^{(s)}(k))}{d\hat{W}_1^{(s)}(k)}^T,\ \frac{d\rho_{2w}(\hat{W}_2^{(s)}(k))}{d\hat{W}_2^{(s)}(k)}^T,\ \cdots,\ \frac{d\rho_{Nw}(\hat{W}_N^{(s)}(k))}{d\hat{W}_N^{(s)}(k)}^T]^T \tag{4.72a}$$

$$\frac{d\rho_{iw}(\hat{W}_i^{(s)}(k))}{d\hat{W}_i^{(s)}(k)} = [\frac{d\rho(\hat{w}_1^{(s)}(k-H_p+1))}{d\hat{w}_1^{(s)}(k-Hp+1)}^T,\ \frac{d\rho(\hat{w}_2^{(s)}(k-H_p+2))}{d\hat{w}_2^{(s)}(k-H_p+2)}^T,\ \cdots,\ \frac{d\rho(\hat{w}_{ni}^{(s)}(k-1))}{d\hat{w}_{ni}^{(s)}(k-1)}^T]^T \tag{4.72b}$$

$$\frac{d\rho(\hat{w}_i^{(s)}(k-H_p+1))}{d\hat{w}_i^{(s)}(k-Hp+1)} = [\rho'(\hat{w}_{i1}^{(s)}),\ \rho'(\hat{w}_{i2}^{(s)}),\ \cdots,\ \rho'(\hat{w}_{in_i}^{(s)})]^T \tag{4.72c}$$

Since the objective functions in all three schemes (centralized MHE, decentralized MHE and prediction-driven CDMHE) are all minimization problems, after adding the barrier function to the objective function, the barrier function assures that the variables satisfy the constraints. Compared with the unconstrained case, the optimum that the augmented centralized MHE find is inside the boundary and may be different form the unconstrained case. For the augmented prediction-driven CDMHE , for every iteration at each sampling time, there exists sequences of $\hat{X}^{(s)}$ and $\hat{W}^{(s)}$ inside the boundary.

In the implementation of the CDMHE algorithm, the basic idea is to approximate the barrier function defined in (4.64). Let us consider the case that we only have the upper bound $u \leq 0$. The barrier function denoted as $\rho_{-}(u)$ becomes:

$$\rho_{-}(u) = \begin{cases} 0 & u \leq 0 \\ +\infty & else \end{cases} \tag{4.73}$$

Logarithmic barrier function is one way to approximate the barrier function (4.64). As shown in Figure 4.2, logarithmic barrier functions give good approximations. In the implementation, the approximation of the barrier function always ensures the variables are inside the boundary. The logarithmic function that used to approximate the barrier function (4.64) is expressed as:

$$\hat{\rho}_{-}(u) = -(1/t)log(-u) \tag{4.74}$$

where $-(1/t)log(-u)$ is convex and increasing in u, and differentiable. $t > 0$ is a parameter that sets the accuracy of the approximation. As t increases, the approximation becomes more accurate (Boyd and Vandenberghe, 2004).

4.3.1 Convergence Analysis

Using the logarithmic barrier function to approximate the prediction-driven CDMHE inequality constraints (4.63). For the aggregated augmented prediction-driven CDMHE and the augmented centralized MHE, denoting the logarithmic barrier function in the cost function as $(1/t)\phi(X)$, it can be obtained that:

$$\mathcal{J}_{CDMHEC} = \mathcal{J}_{CDMHE} + (1/t)\phi(X) \tag{4.75a}$$

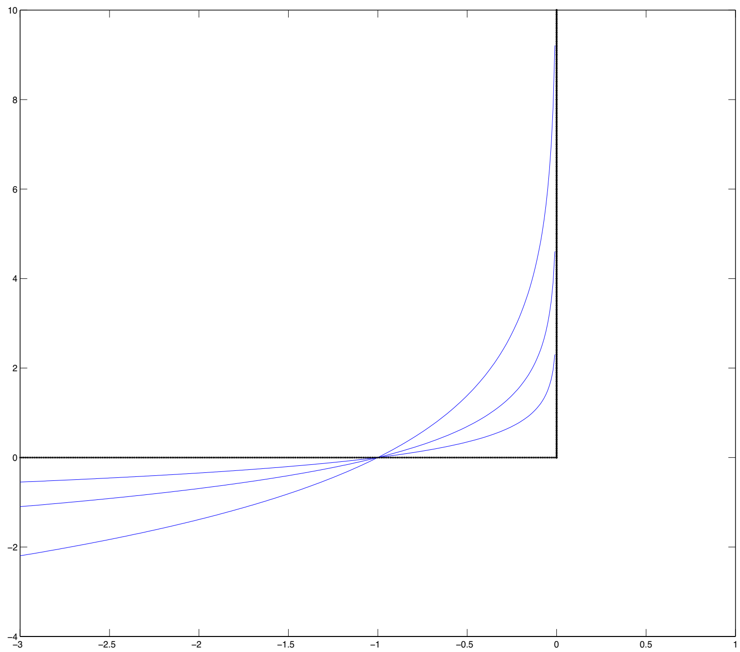

Figure 4.2: The dashed lines show the function $\rho_{-}(u)$ and the solid lines show $\hat{\rho}_{-}(u) = -(1/t)log(-u)$.

$$\mathcal{J}_{cenc} = \mathcal{J}_{cen} + (1/t)\phi(X) \tag{4.75b}$$

Let us consider that the system has m inequality constraints that can be expressed as $T_i(X) \leq 0$, for $i = 1, \ \ldots, \ m$; therefore, $\phi(X) = -\sum_{i=1}^{m} log(-T_i(X))$. The first order and second order derivatives of logarithmic barrier function ϕ can be expressed as (Boyd and Vandenberghe, 2004):

$$\nabla\phi(X) = \sum_{i=1}^{m} \frac{1}{-T_i(X)} \nabla T_i(X) \tag{4.76a}$$

$$\nabla^2\phi(X) = \sum_{i=1}^{m} \frac{1}{T_i(X)^2} \nabla T_i(X) \nabla T_i(X)^T + \sum_{i=1}^{m} \frac{1}{-T_i(X)} \nabla^2 T_i(X) \tag{4.76b}$$

Therefore, for the augmented centralized MHE, equation (4.54) becomes:

$$\frac{d\mathcal{J}_{cenc}}{d\hat{X}(k)}|_{\hat{X}^{(s)}(k)} = (\Psi + \frac{1}{t}\nabla^2\phi(X^*))(\hat{X}^{(s)}(k) - \hat{X}^*(k)) \tag{4.77}$$

For the augmented prediction-driven CDMHE, equation (4.58) becomes:

$$\frac{d\mathcal{J}_{CDMHEC}}{d\hat{X}(k)}|_{\hat{X}^{(s)}(k)} = (\Omega + \frac{1}{t}\nabla^2\phi(X^{(s+1)}))(\hat{X}^{(s)}(k) - \hat{X}^{(s+1)}(k)) \tag{4.78}$$

For the derivatives of the logarithmic barrier function $\phi(X)$, we have the following assumption:

- With the increasing of t to ∞, $\frac{1}{t}\nabla\phi(X)$ and $\frac{1}{t}\nabla^2\phi(X)$ are going to zero.

Let us used the following symbols to represent the derivatives of logarithmic barrier function as:

$$\varepsilon_1 = \frac{1}{t}\nabla^2\phi(X^*)(\hat{X}^{(s)}(k) - \hat{X}^*(k)), \tag{4.79a}$$
$$\varepsilon_2 = \frac{1}{t}\nabla^2\phi(X^{(s+1)})(\hat{X}^{(s)}(k) - \hat{X}^{(s+1)}(k)) \tag{4.79b}$$
$$\varepsilon_3 = \frac{1}{t}\nabla^2\phi(X^{(s)}) \tag{4.79c}$$

A similar equality equation to (4.59) can be obtained as:

$$\begin{aligned}\hat{X}^{(s)}(k) - \hat{X}^{(s+1)}(k) &= \Omega^{-1}(\Psi(\hat{X}^{(s)}(k) - \hat{X}^*(k)) + \varepsilon_1 - \varepsilon_2) \\ &= \Omega^{-1}\Psi(\hat{X}^{(s)}(k) - \hat{X}^*(k)) + \varepsilon_4\end{aligned} \tag{4.80}$$

where $\varepsilon_4 = \Omega^{-1}(\varepsilon_1 - \varepsilon_2)$.

Therefore, for the augmented cost functions, the similar expression to (4.62) can be obtained as following:

$$\begin{aligned}\mathcal{J}_{cenc}(\hat{X}^{(s)}(k)) - \mathcal{J}_{cenc}(\hat{X}^{(s+1)}(k)) =&(\Omega^{-1}\Psi(\hat{X}^{(s)}(k) - \hat{X}^*(k)) + \varepsilon_3)^T(\Omega - \frac{\Psi}{2} - \frac{\varepsilon_4}{2}) \\ &(\Omega^{-1}\Psi(\hat{X}^{(s)}(k) - \hat{X}^*(k)) + \varepsilon_3)\end{aligned} \tag{4.81}$$

The values of $\varepsilon_1,\ \varepsilon_2,\ \varepsilon_3$ and ε_4 are closely related to the value of t which can be adjusted in the simulations. According to the assumption that when t goes to ∞, $\epsilon_i,\ i = 1,\ 2,\ 3,\ 4$ are all close to zero. Therefore the similar conclusion can be obtained as in Section 4.2.2 as t goes to ∞, with the increasing of iteration s, the estimate obtained by the augmented prediction-driven CDMHE $\hat{X}^{(s)}(k)$ converges to the augmented centralized estimate $\hat{X}^*(k)$.

4.4 Illustrative Examples

In this section, the proposed prediction-driven CDMHE algorithm is applied to two chemical processes. In the first case, the process contains two connected continuous stirred tank reactor. In the second case, the process consists of two connected continuous stirred tank reactors and one flash tank separator which has been used in Section 3.5. In this case, we consider constraints on the estimated variables $\hat{X}$ and $\hat{W}$ and use the method proposed in Section 4.3 to transfer the inequality constraints into the objective function. The results of both cases show that the proposed prediction-driven CDMHE reaches the centralized MHE performance.

4.4.1 Two-CSTR Case

In this section, the proposed coordinated state estimation approach is applied to a simulated chemical process. The process contains two connected continuous stirred tank reactor (CSTR) with a recycle stream as shown in Figure 4.3 (Sun and El-Farra, 2008). As shown in Figure 4.3, pure A is fed into the first CSTR at flow rate F_0, molar concentration C_{A0} and temperature T_0. A stream recycled from CSTR 2 is also fed into CSTR 1 at flow rate F_r, molar concentration C_{A2} and temperature T_2. The outlet stream of CSTR 1 is fed to CSTR 2 and an additional stream containing pure A at molar concentration C_{A03}, flow rate F_3 and temperature T_{03} is also fed to CSTR 2. The output of CSTR 2 passes through a separator which is used to remove the products and recycle unreactant A back to CSTR 1. Three irreversible elementary exothermic reactions $A \to B$, $A \to U$ and $A \to R$ take place in the two reactors, where A is the reactant material, B is the desired product, and R and U are the undesired byproducts. Because of the non-isothermal nature of the reactions, each CSTR is equipped with a jacket to remove/provide heat. Based on standard modeling assumptions, four ordinary differential equations can be obtained to describe the dynamics:

$$\frac{dT_1}{dt} = \frac{F_0}{V_1}(T_0 - T_1) + \frac{F_r}{V_1}(T_2 - T_1) + \sum_{i=1}^{3} G_i(T_1)C_{A1} + \frac{Q_1}{\rho c_p V1} \tag{4.82a}$$

$$\frac{dC_{A1}}{dt} = \frac{F_0}{V_1}(C_{A0} - C_{A1}) + \frac{F_r}{V_1}(C_{A2} - C_{A1}) - \sum_{i=1}^{3} R_i(T_1)C_{A1} \tag{4.82b}$$

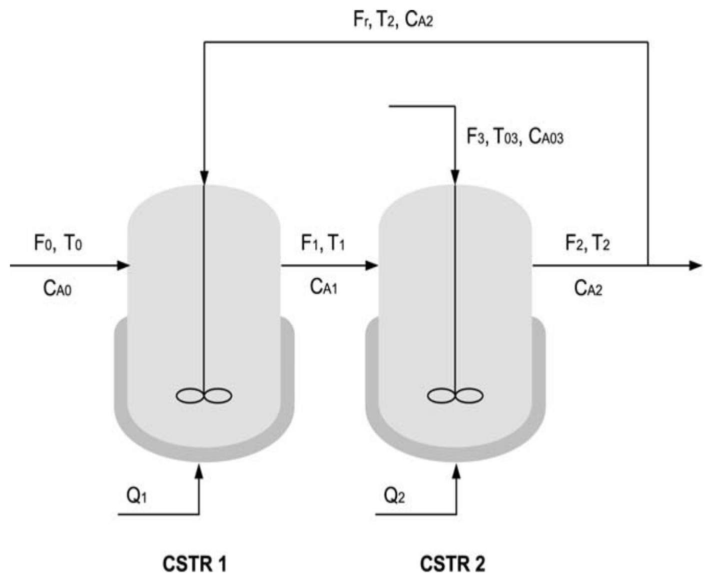

Figure 4.3: Process flow diagram of two interconnected CSTR units.

$$\frac{dT_2}{dt} = \frac{F_1}{V_2}(T_1 - T_2) + \frac{F_3}{V_2}(T_{03} - T_2) + \sum_{i=1}^{3} G_i(T_2)C_{A2} + \frac{Q_2}{\rho c_p V2} \tag{4.82c}$$

$$\frac{dC_{A2}}{dt} = \frac{F_1}{V_2}(C_{A1} - C_{A2}) + \frac{F_3}{V_2}(C_{A03} - C_{A2}) - \sum_{i=1}^{3} R_i(T_2)C_{A2} \tag{4.82d}$$

where $R_i(T_j) = k_{i0}exp(-E_i/RT_j)$, $G_i(T_j) = ((-\Delta H_i)/\rho c_p)R_i(T_j)$, with $j = 1, 2$, T_j, C_{Aj}, V_j and Q_j represent the temperature of the reactor, the concentration of A, the reactor volume and the rate of the heat input to the reactor, respectively, $\Delta H_i, k_i, E_i$, with $i = 1, 2, 3$, denote the enthalpies, pre-exponetial constants and activation energies of three reactions, respectively, c_p and ρ denote the heat capacity and density of the fluid in the reactors. The parameters are given in Table 3.2. When $Q_1 = Q_2 = 0$, $C_{A0} = C_{A0}^s$, $C_{A03} = C_{A03}^s$, the process has a steady state at $(T_1^s, C_{A1}^s, T_2^s, C_{A2}^s) = (300.3878K, 2.4881kmol/m^3, 300.3496K, 2.2840kmol/m^3)$. Linearizing the nonlinear model at this steady state, a linear model of the process can be obtained. We assume that temperatures T_1, T_2 are measurable and sampled synchronously with sampling time $\Delta t = 0.005h = 18sec$. The concentration in the two CSTRs are unmeasurable and should be estimated. Note that an output feedback controller is implemented to ensure the closed-loop stability. The entire process is divided into two subsystems according to the two reactors.

For subsystem 1, $x_1(k) = [T_1(k) - T_1^s, C_{A1}(k) - C_{A1}^s]^T$, for subsystem 2, $x_2(k) =$

Table 4.1: Process parameters of 2-CSTR process

$F_0 = 4.998m^3/h$	$k_{10} = 3.0 \times 10^6 h^{-1}$
$F_1 = 39.996m^3/h$	$k_{20} = 3.0 \times 10^5 h^{-1}$
$F_3 = 30.0m^3/h$	$k_{30} = 3.0 \times 10^5 h^{-1}$
$F_r = 34.998m^3/h$	$E_1 = 5.0 \times 10^4 kJ/kmol$
$V_1 = 1.0m^3$	$E_1 = 7.53 \times 10^4 kJ/kmol$
$V_1 = 3.0m^3$	$E_1 = 7.53 \times 10^4 kJ/kmol$
$R = 8.314kJ/kmolK$	$\rho = 1000.0kg/m^3$
$T_0 = 300.0K$	$c_p = 0.231kJ/kgK$
$T_{03} = 300.0K$	$C_{A0}^s = 4.0kmol/m^3$
$C_{A03}^s = 2.0kmol/m^3$	$\Delta H_1 = -5.0 \times 10^4 kJ/kmol$
$\Delta H_2 = -5.2 \times 10^4 kJ/kmol$	$\Delta H_3 = -5.4 \times 10^4 kJ/kmol$

Table 4.2: Parameters used in the prediction-driven CDMHE for the 2-CSTR process

	MHE 1	MHE 2
Initial Guess	$\hat{x}_1(0) = [5; 3.5]$	$\hat{x}_2(0) = [20; 4]$
Moving Horizon	10	10
Weighting Matrix	$Q_1 = \begin{bmatrix} 0.6^2 & 0 \\ 0 & 0.98^2 \end{bmatrix}$	$Q_2 = \begin{bmatrix} 0.6^2 & 0 \\ 0 & 0.98^2 \end{bmatrix}$
	$R_1 = 0.8^2$	$R_2 = 0.8^2$

$[T_2(k) - T_2^s, C_{A2}(k) - C_{A2}^s]^T$. The actual initial condition for the whole system is $x(0) = [13,\ 2,\ 34,\ 1.5]^T$. The random disturbances added to the dynamics of the temperatures are generated as normally distributed values with zero mean and standard deviation 1. The random disturbances added to the dynamics of the concentrations are generated as normally distributed values with zero mean and standard deviation 0.1. The initial guesses, horizon and weighting matrices used in the two local MHEs are shown in Table 4.2. The termination threshold $\epsilon = 0.0001$.

For this chemical process, there are two subsystems. According to (4.38), we can get Ψ as follows:

$$\Psi = \mathbb{C}^T \boldsymbol{R}^{-1} \mathbb{C} + G_A^T G_B^{-1^T} \boldsymbol{Q}^{-1} G_B^{-1} G_A \tag{4.83}$$

where $\mathbb{C}, \boldsymbol{R}$ and $\boldsymbol{Q}$ are defined in (2.6), G_A and G_B are defined in (2.13).

According to equation (4.38), we can get Ω as follows:

$$
\begin{aligned}
\Omega_1 &= \boldsymbol{C}_{11}^T \boldsymbol{R}_1^{-1} \boldsymbol{C}_{11} + G_{A_{11}}^T {G_{B_{11}}^{-1}}^T \boldsymbol{Q}_1^{-1} G_{B_{11}}^{-1} G_{A_{11}} \\
\Omega_2 &= \boldsymbol{C}_{22}^T \boldsymbol{R}_2^{-1} \boldsymbol{C}_{22} + G_{A_{22}}^T {G_{B_{22}}^{-1}}^T \boldsymbol{Q}_2^{-1} G_{B_{22}}^{-1} G_{A_{22}} \\
\Omega &= \begin{bmatrix} \Omega_1 & \\ & \Omega_2 \end{bmatrix}
\end{aligned}
\tag{4.84}
$$

It can be verified that the eigenvalues of $(\Omega - \Psi/2)$ are all positive which means $(\Omega - \Psi/2)$ is positive definite, so the convergence condition proposed in **Proposition** 4.2.1 is satisfied.

The coordinated scheme is implemented following Algorithm 2 to estimate the entire system state in a distributed way. The proposed prediction-driven CDMHE is compared with different estimation techniques to illustrate its performance. Specially, the proposed prediction-driven CDMHE is compared with the corresponding centralized MHE and decentralized MHE. Note that in the decentralized MHE, the interactions between subsystems are neglected.

The simulation results are given in Figure 4.4 - Figure 4.6. The trajectories of the estimates given by the three estimation schemes are shown in Figure 4.4. In order to see the estimated trajectories obtained by the prediction-driven CDMHE algorithm and the centralized MHE clearly. Figure 4.5 shows the estimation given by the centralized MHE and the prediction-driven CDMHE along with the actual state trajectory. The trajectories of the estimation error are in Figure 4.6.

From Figure 4.5, it can be seen that the estimated state given by the proposed prediction-driven CDMHE and the estimated states given by the centralized MHE are very close (the two lines overlap with each other). This means the prediction-driven CDMHE has reached the performance of the centralized MHE. From Figure 4.4 and Figure 4.6, we can see that the estimation given by the decentralized MHE is poor compared with the prediction-driven CDMHE. Note that both the centralized MHE and the prediction-driven CDMHE give relatively poor estimates of the second state. This is due to the characteristics of the process. In this work, we show that the proposed prediction-driven CDMHE gives the same performance as the centralized MHE. From Figure 4.9, it can be seen that the proposed prediction-driven CDMHE and the centralized MHE give significantly improved performance over the decentralized MHE. This is because the prediction-driven CDMHE and centralized MHE take

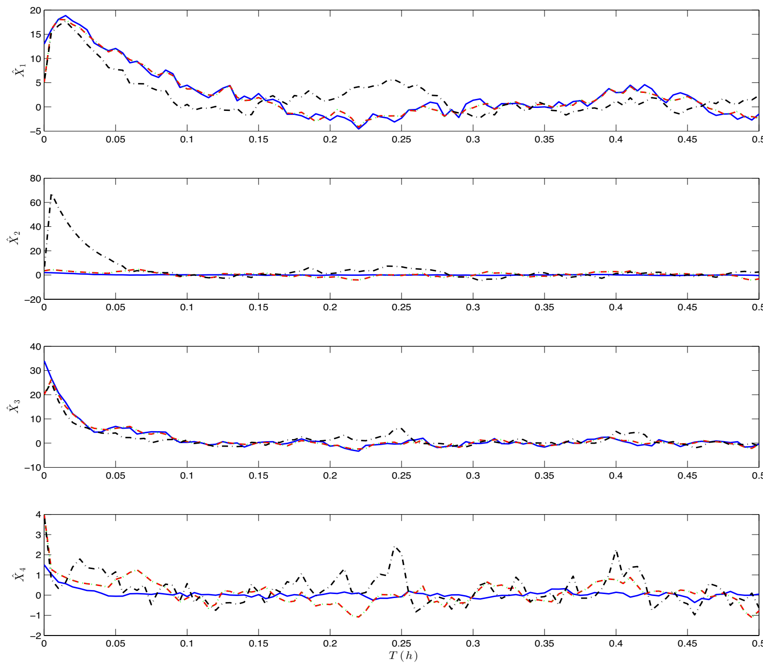

Figure 4.4: Trajectories of the actual state (solid line), estimates given by the proposed prediction-driven CDMHE (dashed line), estimates given by the centralized MHE (dotted line), and the decentralized MHE (dash-dotted line).

into account the interactions between subsystem explicitly while in the decentralized MHE the interaction between subsystems is ignored.

4.4.2 Two-CSTR and One Separator

The same chemical process used in Section 3.5 is used here to illustrate the effectiveness of the prediction-driven CDMHE. The initial state remains same. We use logarithmic barrier to approximate the barrier functions mentioned in (4.64). For subsystem i in the prediction-driven CDMHE and decentralized MHE, the barrier functions are chosen as:

$$\rho_{ix}(\hat{X}_i(k)) = (-1/t) \sum_{l=K-H_p+1}^{K} \sum_{j=1}^{n_i} (log(\hat{x}_{ij}(l) - \mathbb{X}lb_{ij}) + log(\mathbb{X}ub_{ij} - \hat{x}_{ij}(l)) \tag{4.85a}$$

$$\rho_{iw}(\hat{W}_i(k)) = (-1/t) \sum_{l=K-H_p+1}^{K-1} \sum_{j=1}^{n_i} (log(\hat{w}_{ij}(l) - \mathbb{W}lb_{ij}) + log(\mathbb{W}ub_{ij} - \hat{x}_{ij}(l)) \tag{4.85b}$$

In the centralized MHE, the barrier function is the summation of the decentralized

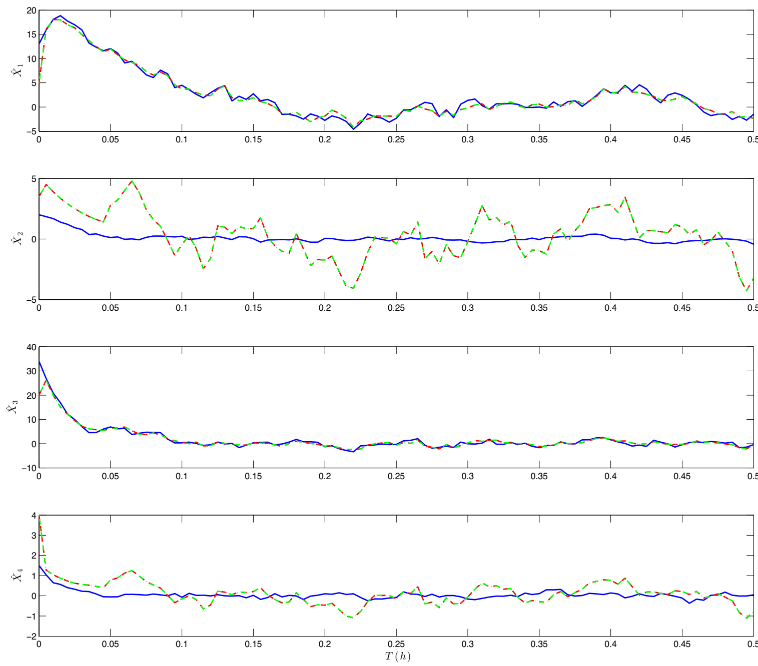

Figure 4.5: Trajectories of the actual state (solid line), estimates given by the proposed prediction-driven CDMHE (dashed line), estimates given by the centralized MHE (dotted line).

ones, as following:

$$\rho_x(\hat{X}(k)) = \sum_{i=1}^{N} \rho_{ix}(\hat{X}_i(k)) \tag{4.86a}$$

$$\rho_w(\hat{W}(k)) = \sum_{i=1}^{N} \rho_{iw}(\hat{W}_i(k)) \tag{4.86b}$$

where t is a turning parameter, $\mathbb{X}lb$ and $\mathbb{X}ub$ are the lower bound and upper bound of the constraints on the estimated states $\hat{X}$, respectively, and $\mathbb{W}lb$ and $\mathbb{W}ub$ are the lower bound and upper bound of the constraints on the estimated noises $\hat{W}$, respectively.

The tuning parameter t is chosen as a large positive number and determines the accuracy of the approximation. When $\hat{x}_{ij}$ and $\hat{w}_{ij}$ are inside the boundary, the barrier function terms will almost have no effects on the original cost function. When $\hat{x}_{ij}$ or $\hat{w}_{ij}$ approaches the boundary, the value of $\rho_{xi}(\hat{X}_i(k))$ or $\rho_{wi}(\hat{W}_i(k))$ approaches $+\infty$. In this case, the tuning parameter t is chosen as 10^6. The weighting matrices and initial guesses of subsystems are given in Table 4.3.

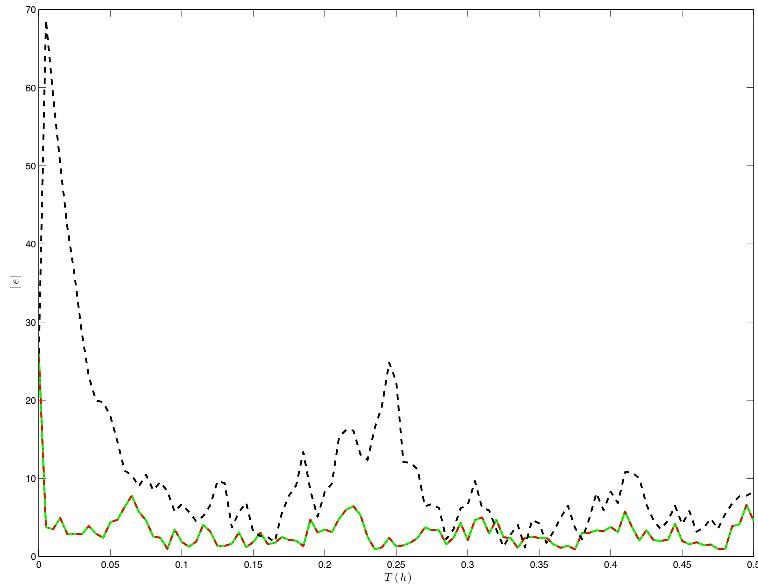

Figure 4.6: Trajectories of the estimation error given by proposed prediction-driven CDMHE (dash-dotted line), the centralized MHE (solid line), and the decentralized MHE (dashed line)

For this chemical process, there are three subsystems. According to (4.38), we can get Ψ as follows:

$$\Psi = \mathbb{C}^T \boldsymbol{R}^{-1} \mathbb{C} + G_A^T {G_B^{-1}}^T \boldsymbol{Q}^{-1} G_B^{-1} G_A \tag{4.87}$$

where $\mathbb{C}, \boldsymbol{R}$ and $\boldsymbol{Q}$ are defined in (2.6), G_A and G_B are defined in (2.13).

Table 4.3: Parameters used in the prediction-driven CDMHE

	Initial Guess	Moving Horizon	Weighting Matrix
MHE 1	$\hat{x}_1(0) = [0.0508\ 0.3277\ 561.0186]^T$	10	$Q_1 = \begin{bmatrix} 0.82^2 & 0 & \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.8^2 \end{bmatrix}$ $R_1 = 1^2$
MHE 2	$\hat{x}_2(0) = [0.1518\ 0.4444\ 513.3340]^T$	10	$Q_2 = \begin{bmatrix} 0.52^2 & 0 & 0 \\ 0 & 0.8^2 & 0 \\ 0 & 0 & 1.2^2 \end{bmatrix}$ $R_2 = 1^2$
MHE 3	$\hat{x}_3(0) = [0.0693\ 0.3560\ 546.4699]^T$	10	$Q_3 = \begin{bmatrix} 0.2^2 & 0 & 0 \\ 0 & 0.1^2 & 0 \\ 0 & 0 & 0.3^2 \end{bmatrix}$ $R_3 = 1^2$

According to equation (4.38), we can get Ω as follows:

$$\begin{aligned}
\Omega_1 &= \boldsymbol{C}_{11}^T \boldsymbol{R}_1^{-1} \boldsymbol{C}_{11} + G_{A_{11}}^T {G_{B_{11}}^{-1}}^T \boldsymbol{Q}_1^{-1} G_{B_{11}}^{-1} G_{A_{11}} \\
\Omega_2 &= \boldsymbol{C}_{22}^T \boldsymbol{R}_2^{-1} \boldsymbol{C}_{22} + G_{A_{22}}^T {G_{B_{22}}^{-1}}^T \boldsymbol{Q}_2^{-1} G_{B_{22}}^{-1} G_{A_{22}} \\
\Omega_3 &= \boldsymbol{C}_{33}^T \boldsymbol{R}_3^{-1} \boldsymbol{C}_{33} + G_{A_{33}}^T {G_{B_{33}}^{-1}}^T \boldsymbol{Q}_3^{-1} G_{B_{33}}^{-1} G_{A_{33}} \\
\Omega &= \begin{bmatrix} \Omega_1 & & \\ & \Omega_2 & \\ & & \Omega_3 \end{bmatrix}
\end{aligned} \tag{4.88}$$

It can be verified that the eigenvalues of $(\Omega - \Psi/2)$ are all positive which means $(\Omega - \Psi/2)$ is positive definite, so the convergence condition proposed in **Proposition** 4.2.1 is satisfied.

The simulation results are shown in Figure 4.7 - Figure 4.10. Figure 4.7 shows the trajectories of the estimated states given by the centralized MHE, the decentralized MHE, the prediction-driven CDMHE and the actual states.Figure 4.8 is used to give us a clearer view of the estimated states given by the prediction-driven CDMHE and the centralized MHE. We can see that the prediction-driven CDMHE reaches the centralized performance. Figure 4.9 shows the trajectories of the error norm of the three estimated algorithms. We can see that the performance of centralized MHE and proposed prediction-driven CDMHE algorithm are much better than the decentralized MHE. The summation of error given by the prediction-driven CDMHE is 14.0238 while the summation of error given by the decentralized MHE is 147.15 which is more than 10 times of the error norm provided by prediction-driven CDMHE. Figure 4.10 shows the number of iterations during the sampling time intervals. It can be seen that the iterations are between 8 and 22.

From the above two simulation cases, the performance of proposed prediction-driven CDMHE agrees with what is suggested by the theoretical analysis, and reaches the performance of the centralized MHE. Compared with the decentralized MHE, the proposed prediction-driven CDMHE has significant improvement while still keeps the flexibility in the structure.

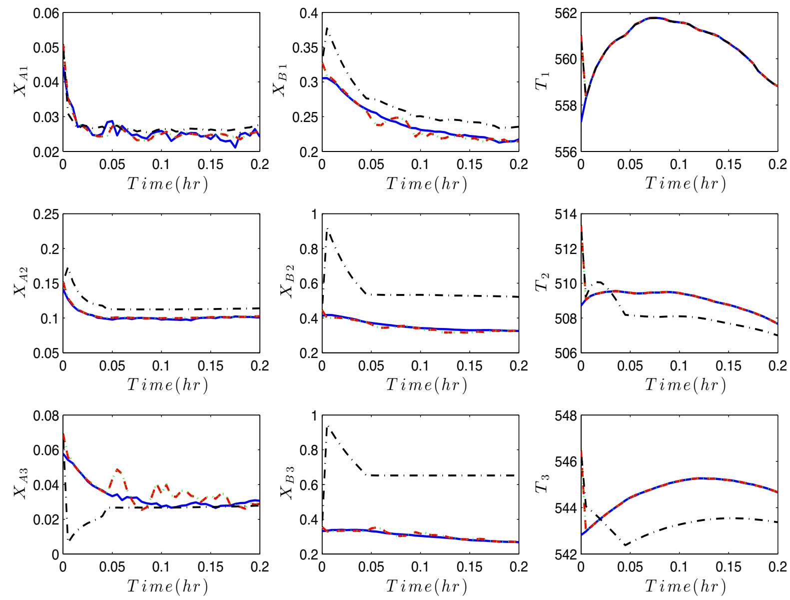

Figure 4.7: Trajectories of the actual state (solid line), estimates given by the proposed CDMHE (dashed line), estimates given by the centralized MHE (dotted line), and the decentralized MHE (dash-dotted line).

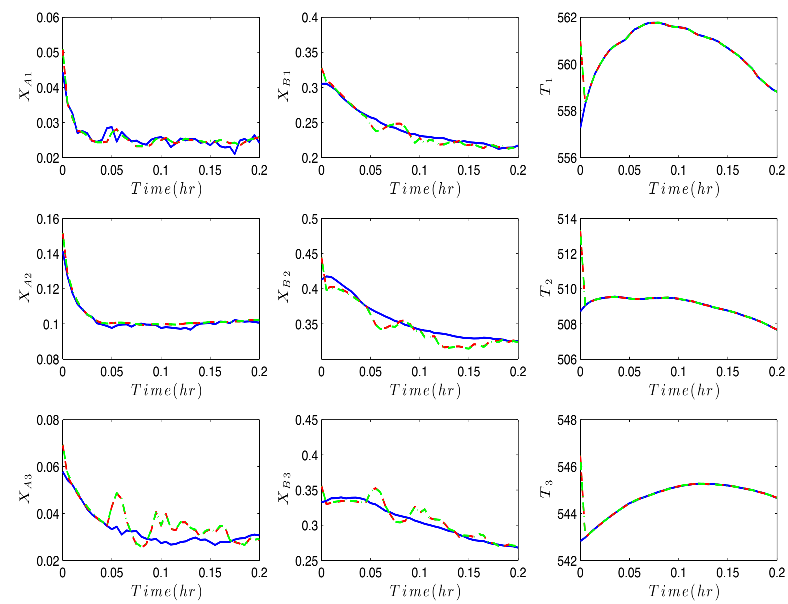

Figure 4.8: Trajectories of the actual state (solid line), estimates given by the proposed CDMHE (dashed line), estimates given by the centralized MHE (dash-dotted line).

4.5 Conclusions

In this chapter, a coordinated distributed moving horizon estimation scheme is proposed for a class of discrete-time, linear systems. In particular, the class of linear

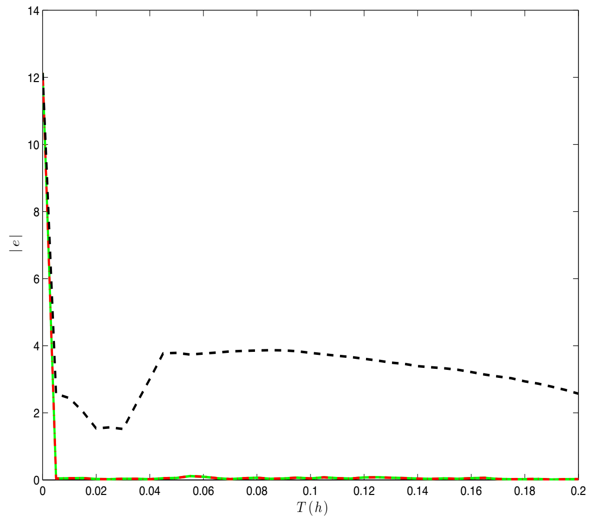

Figure 4.9: Trajectories of the estimation error norm given by proposed CDMHE (dash-dotted line), the centralized MHE (solid line), and the decentralized MHE (dashed line).

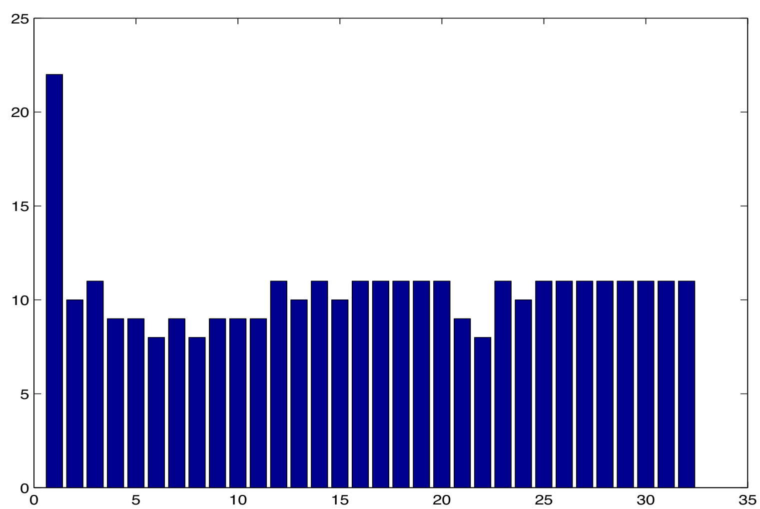

Figure 4.10: Numbers of iterations during each sampling time.

systems we focus on is composed of several subsystems that interact with each other via their states. It is seen from the formulations that the proposed prediction-driven CDMHE estimators are obtained by modifying the decentralized MHE estimator. A coordinating term is added to each local MHE to connect the local estimators with the coordinator. After modifying the decentralized MHEs, the coordinator is designed based on a method used to solve the the optimality conditions of the aggregated CDMHE. This allows the prediction-driven CDMHE scheme to maintain the desired properties of the decentralized MHE estimators such as flexibility and robustness.

Convergence of the prediction-driven CDMHE algorithm is also studied in this chapter and sufficient convergence conditions are given. In addition, a way to use barrier function to handle inequality constraints in the proposed coordinated algorithm is presented in Section 4.3 and the convergence is guaranteed. Furthermore, it is shown that once the convergence condition is satisfied, the solution of the proposed prediction-driven CDMHE algorithm converges to the centralized MHE solution. Two chemical processes are used to illustrate the efficiency and applicability of the proposed coordinated scheme.

Chapter 5

Robustness of the Prediction-driven CDMHE

In this chapter, the robustness of the prediction-driven CDMHE algorithm proposed in Chapter 4 is investigated under different scenarios. These scenarios include: 1) Triggered communication, between the subsystem estimators and the coordinator, 2) Communication failure between a subsystem and the coordinator, and 3) Premature termination of the coordination algorithm. The robustness investigation is carried out based primarily on simulations with extensive discussion.

5.1 Triggered Communication

In this section, the proposed prediction-driven CDMHE with triggered communication is discussed. The purpose of the triggered communication method is to minimize the communication cost between the local estimators and the coordinator. A schematic of the prediction-driven CDMHE with triggered communication is presented in Figure 5.1. In this algorithm, every local subsystem has a MHE estimator and a communication trigger which determines whether the information in the current iteration should be sent to the coordinator. This implies that the information obtained in a subsystem estimator at the current iteration is not necessarily sent to the coordinator. In this way the communication load between the subsystems and the coordinator can be reduced. In this algorithm, the difference between the current state estimate and the last sent state estimate is the basis to design the triggering condition. The rest of this section is organized as follows: first, the proposed prediction-driven CDMHE with triggered communication algorithm is presented; then a convergence analysis is given; finally, the algorithm is applied to the chemical process that described in Section 3.5.

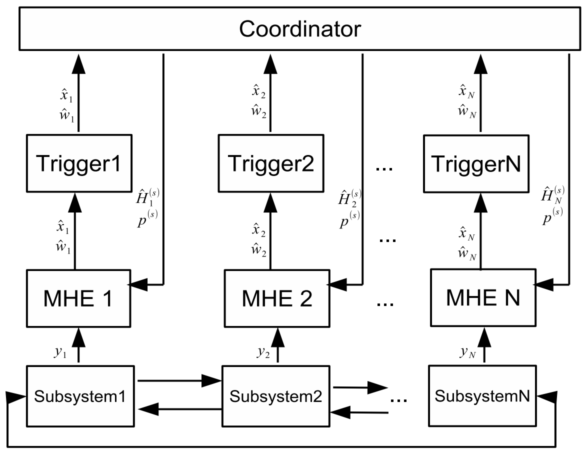

Figure 5.1: A schematic of the proposed prediction-driven CDMHE design with triggered communication.

The implementation algorithm of the prediction-driven CDMHE with triggered communication is described in Algorithm 3.

It is shown in Algorithm 3 that the triggering condition for each subsystem i is checked every iteration during each sampling interval after the latest estimated state is calculated by the local MHE i. The triggering condition is designed based on the difference between the current estimated states and last sent estimated states. For subsystem i, the triggering condition at s^{th} iteration during sampling time k is defined specifically as follows:

$$U_i^{(s)}(k) = \begin{cases} 1, & \text{if } \|\hat{X}_{i(s)}^*(k) - \hat{X}_i^{(s_i)}(k)\| \geq \epsilon_i \\ 0, & \text{if } \|\hat{X}_{i(s)}^*(k) - \hat{X}_i^{(s_i)}(k)\| < \epsilon_i \end{cases} \tag{5.1}$$

where s_i is the last iteration that MHE i sent its state to the coordinator. $\hat{X}_{i(s)}^*(k)$ and $\hat{W}_{i(s)}^*(k)$ are the current estimated state and noise of MHE i at iteration s in the time interval k, respectively. $\hat{X}_i^{(s_i)}(k)$ and $\hat{W}_i^{(s_i)}(k)$ are the last sent estimated state and noise of MHE i at time interval k. ϵ_i is the threshold of subsystem i. When $U_i^{(s)}(k) = 1$, that is, the triggering condition is satisfied, the MHE i set $\hat{X}_i^{(s)}(k) = \hat{X}_{i(s)}^*(k)$, $\hat{W}_i^{(s)}(k) = \hat{W}_{i(s)}^*(k)$ and sends them to the coordinator. At the same time, the local trigger updates s_i to s. When $U_i^{(s)}(k) = 0$, which implies the triggering

Algorithm 3 Prediction-driven CDMHE with triggered communication

Initialization

Coordinator : Set iteration counter $s = 0$. When $k = 1$, $\hat{X}^{(0)}(k)$ and $\hat{W}^{(0)}(k)$ are arbitrarily determined; else $\hat{X}^{(0)}(k)$ and $\hat{W}^{(0)}(k)$ are picked as $\hat{X}(k-1)$ and $\hat{W}(k-1)$, and $\boldsymbol{p}^{(0)}(k)$ and $\hat{H}_i^{(s)}(k)$ are calculated according to (4.9) and (4.6), respectively.

Local Trigger : In local trigger i, $\hat{X}_i^{(0)}(k)$ is picked as the corresponding part in $\hat{X}^{(0)}(k)$ in the coordinator.

repeat

Coordinator : $\boldsymbol{p}^{(s)}$ and $\hat{H}_i^{(s)}(k)$ are sent to local MHE estimators.

Local Estimators : Local problem (4.2) is solved based on the local measurements $Y_i(k)$. $\hat{X}^*_{i(s)}(k)$ is sent to the trigger that associated with the local MHE.

Local Trigger : The trigger checks the triggering condition. If the triggering condition is satisfied, the trigger updates $\hat{X}_i^{(s)}(k) = \hat{X}^*_{i(s)}(k)$ and $\hat{W}_i^{(s)}(k) = \hat{W}^*_{i(s)}(k)$ and sends $\hat{X}_i^{(s)}(k)$ and $\hat{W}_i^{(s)}(k)$ to the coordinator. Otherwise, no information is transmitted between the trigger and the coordinator.

Coordinator : Coordinator uses $\hat{X}_i^{(s)}(k)$ and $\hat{W}_i^{(s)}(k)$ that received from triggers to construct $\hat{X}^{(s)}(k)$ and $\hat{W}^{(s)}(k)$. Then, $\boldsymbol{p}^{(s)}$ and $\hat{H}_i^{(s)}(k)$ are calculated based on (4.9) and (4.6).

until stopping criterion $\|\hat{X}^{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \epsilon$ is satisfied.

condition is not satisfied, the MHE i does not send any estimated variables to the coordinator. The coordinator continues using the estimated variables that received previously. From the triggering condition in equation (5.1), it can be seen that the condition is dependent on the local subsystem. Therefore, the triggering conditions for different subsystems may be satisfied at different iterations.

From Algorithm 3, it can be seen that a local trigger uses the triggering condition to decide whether to update the estimated state or not. In the prediction-driven CDMHE with triggered communication, instead of using the current estimated states from the subsystems, sometimes coordinator needs to use the last sent estimated states to approximate the current estimated states. The stopping criterion in the prediction-driven CDMHE with triggered communication requires $\|\hat{X}^{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \epsilon$ in the coordinator which is the same as in the prediction-driven CDMHE of Algorithm 2.

5.1.1 Convergence Analysis

The convergence analysis is conducted under the assumption that the proposed prediction-driven CDMHE with regular communication converges. Since the stopping criteria used in the coordinator is $\|\hat{X}^{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \epsilon$, which means the largest difference between two iterations is ϵ. For a system with N subsystems, when the stopping criterion $\|\hat{X}^{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \epsilon$ in the coordinator is satisfied at time k, there are three cases need to be discussed:

1. The triggering conditions and the stopping criteria in the coordinator are all satisfied at the same iteration, which means that all the current estimated states are sent to the coordinator. This situation is exactly the same as the proposed prediction-driven CDMHE with regular communication;

2. Some of the triggering conditions are satisfied. In this case, some of the estimated states $\hat{X}_i^{(s)}(k)$ are last sent estimated states which means they are the approximation of the current actual states $\hat{X}^*_{i(s)}(k)$. Without losing generality, we assume that from subsystem 1 to subsystem l, the triggering conditions are satisfied when the stopping condition in the coordinator is satisfied. Therefore, $\|\hat{X}^*_{i(s)}(k) - \hat{X}_i^{(s-1)}(k)\| < \epsilon_i$ is satisfied in l subsystems. Since $\|\hat{X}^{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \epsilon$, the difference between the current estimated state and last iteration estimated state $\|\hat{X}^*_{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \sqrt{\epsilon^2 + \sum_{i=1}^{l} \epsilon_i^2}$.

3. All the triggering conditions are not satisfied at the end of the iterations at time constant k. In this case, $\hat{X}^{(s)}(k) = \hat{X}^{(s-1)}(k)$ is obtained. Thus, the difference between the current estimated state and last iteration estimated state $\|\hat{X}^*_{(s)}(k) - \hat{X}^{(s-1)}(k)\| < \sqrt{\sum_{i=1}^{N} \epsilon_i^2}$.

From the above three cases, it can be seen that when the iteration ends, the difference between the current estimated state and the last iteration estimated state are bounded within certain threshold. If the proposed prediction-driven CDMHE algorithm converges, the prediction-driven CDMHE with triggered communication will converge as well. It can be seen that the proposed prediction-driven CDMHE with triggered communication gives more relaxed stopping criterion compared with the prediction-driven CDMHE with regular communication. Thus, the communication costs between the distributed MHEs can be reduced by following Algorithm 3 based on the triggering condition in (5.1) with possible bounded loss of estimation performance. Note that by carefully picking the triggering thresholds and the stopping criteria, it is possible to achieve almost the same performance as the prediction-driven CDMHE with regular communication while reducing the communication cost.

5.1.2 Simulation Results

In this section, the proposed prediction-driven CDMHE with triggered communication is compared with the prediction-driven CDMHE with regular communication to illustrate its performance from the communication cost point of view. The chemical process with two-CSTR and one seperator used in Section 4.4.2 is used in this section. Since the stopping criterion used in Section 4.4.2 is $\epsilon = 0.001$, the triggering conditions are picked as $\epsilon_i = 0.001 \times \sqrt{3}/3$, $i = 1,\ 2,\ 3$. The inequality constraints are not considered. Other conditions are the same as the simulation case in Section 4.4.2.

The simulation results are shown in Figures 5.2 - Figures 5.4. Figure 5.2 shows the state trajectories of the estimated sates given by the prediction-driven CDMHE with triggered communication, the centralized MHE and the actual state. From Figure 5.2, it can be seen that the estimated state trajectories given by the prediction-driven CDMHE with triggered communication are almost overlapped with the centralized trajectories. The summation of the absolute error between estimated state given by the prediction-driven CDMHE with triggered communication and the actual state is 14.5753 which is almost same with the summation error term 14.5741 given by the prediction-driven CDMHE with regular communication. Therefore, the prediction-driven CDMHE with triggered communication described in Algorithm 3 keeps the estimation performance of the prediction-driven CDMHE. Figure 5.3 shows the iteration number given by the regular prediction-driven CDMHE and the prediction-driven CDMHE with triggered communication. The iteration number given by the prediction-driven CDMHE with triggered communication is slightly less, which means lower communication cost. Figure 5.4 shows the iteration number that the current estimated state is not sent. From Figure 5.3 and Figure 5.4, it can be concluded that communication cost between the subsystems and the coordinator is reduced.

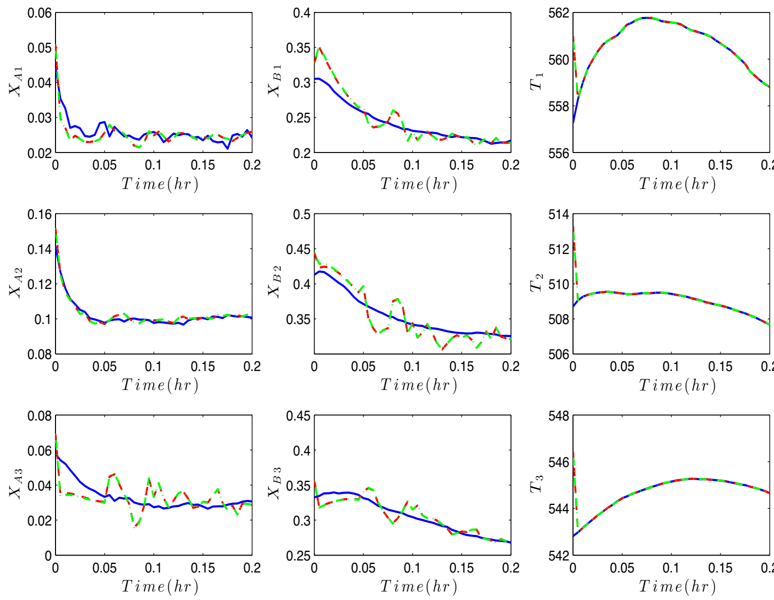

Figure 5.2: State trajectories of the actual states (solid line), the estimated state given by the proposed CDMHE implemented following Algorithm 3 based on triggering condition (5.1) (dashed line) and the estimated state given by the centralized MHE (dash-dotted line).

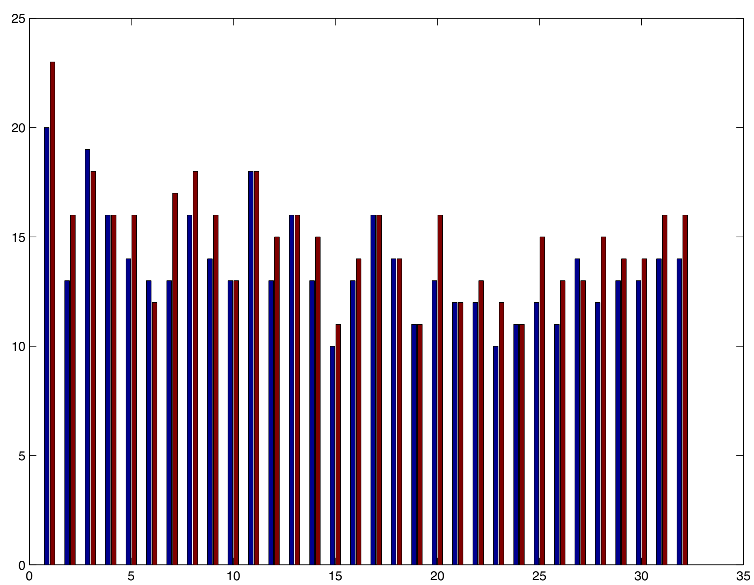

Figure 5.3: Numbers of iterations given by prediction-driven CDMHE (right side) and prediction-driven CDMHE with triggered communication (left side) during each sampling time k.

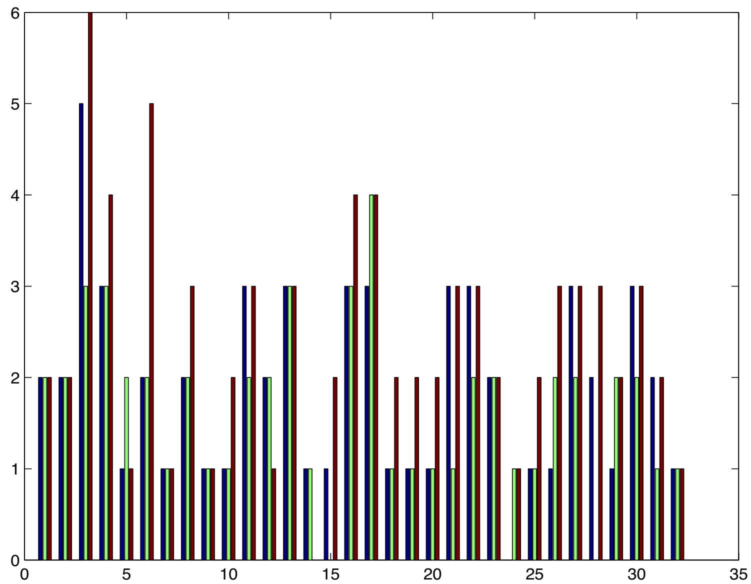

Figure 5.4: Numbers of iterations that subsystem i ($i = 1,\ 2,\ 3$) keeps the last sent estimated state following Algorithm 3 based on triggering condition (5.1) during each sampling time k.

5.2 Communication Failure

Compared with the centralized MHE, the proposed prediction-driven CDMHE maintains the resilience from the decentralized MHE. In the proposed prediction-driven CDMHE, the communication between the coordinator and local MHEs is important in achieving the centralized performance. In this section, the scenario that communication failure happens between subsystems and the coordinator is discussed.

Under the condition that the proposed prediction-driven CDMHE algorithm with regular communication converges, we assume that communication failure happens between subsystem j and the coordinator, which means that coordinator cannot receive any estimated state from subsystem j and subsystem j cannot receive any information about price vector $\boldsymbol{p}$ and interaction $\hat{H}_j$ from the coordinator. There are two cases that need to be discussed:

1. subsystem j loses connection with coordinator from the very beginning and no information has been exchanged between them. Thus, for subsystem j, the price vector $\boldsymbol{p}$ and interactions $\hat{H}_j$ are all zero during all the sampling time, which

makes subsystem j act like the decentralized local MHE. In the coordinator, the calculations of the price vector and interaction vectors will be affected. Therefore, the estimated states of other subsystems, especially the subsystems that strongly coupled with the subsystem j will be affected.

2. when the communication failure happens, subsystem j and the coordinator have exchanged information. In this case, subsystem j keeps using the last received price vector $\boldsymbol{p}$ and interaction vector $\hat{H}_j$ while the coordinator keeps using the last received estimated state from subsystem j. For subsystem j, the price vector $\boldsymbol{p}$ and interaction vector $\hat{H}_j$ are not accurate, the subsystem j will not give good estimated state as before. For other subsystems, the estimation of the price vector and interaction vectors will be affected. Therefore, the estimated states will be affected, especially the subsystems that strongly coupled with the subsystem j.

5.2.1 Simulation Results

In this section, the chemical process with two-CSTR and one seperator introduced in Section 4.4.2 is used. Two cases are discussed. In the first case, subsystem 3 loses connection with coordinator at $k = 0$; in the second case, subsystem 3 loses connection with coordinator after sampling time $k = 15$. The other simulation settings are the same as in Section 4.4.2.

Communication Failure Case 1

In this case, subsystem 3 loses connection with coordinator at $k = 0$. The estimated states should provide by subsystem 3 are picked as zero in the coordinator. Results are shown in Figure 5.5 - Figure 5.7. Figure 5.5 shows the trajectories of estimated states, from which we can see that the estimated state given by subsystem 3 of prediction-driven CDMHE is almost the same with the decentralized MHE. This is caused by the disconnection between the coordinator and subsystem 3. Subsystem 3 cannot receive coordinated variables, and it actually works in the decentralized way. Figure 5.6 gives a clearer view of the trajectories of estimated states given by the prediction-driven CDMHE and the centralized MHE. From Figure 5.6, it can be seen that the estimated state given by the prediction-driven CDMHE is different from those of the centralized MHE. Not only the estimated states of subsystem 3 are affected, the estimated states of other subsystems are also affected. However, the performance of prediction-driven CDMHE is still better than the decentralized MHE which can be seen from the state trajectories in Figure 5.5 and the error norm trajectory in Figure 5.7.

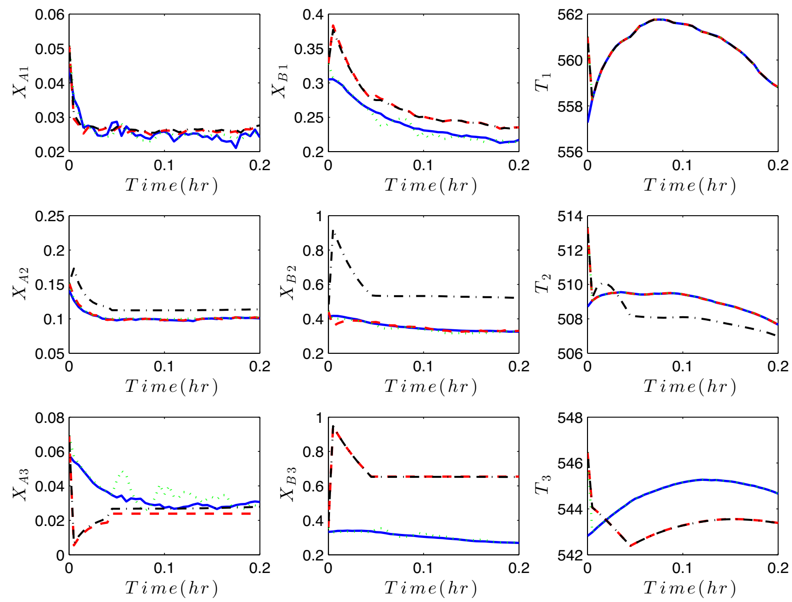

Figure 5.5: Communication Failure Case 1: State trajectories of the actual state (solid line), estimated states given by the proposed CDMHE (dashed line), the centralized MHE (dotted line), and the decentralized MHE (dash-dotted line).

Communication Failure Case 2

In this case, subsystem 3 loses connection with the coordinator after time interval $k = 15$, i.e., $t = 0.075h$. Results are shown in Figure 5.8 - Figure 5.10. Figure 5.8 shows the trajectories of estimated states given by the prediction-driven CDMHE, the centralized MHE and the decentralized MHE, respectively. From Figure 5.9, it can be seen that after $0.075h$, the states of subsystem 3 starts to deviate from the estimated state given by the centralized MHE, other subsystems are also affected. Figure 5.10 shows the trajectories of the error norm of the three estimation algorithms, from which we can see that the error trajectory given by the proposed prediction-driven CDMHE is slightly different from the centralized MHE after $0.075h$. The performance

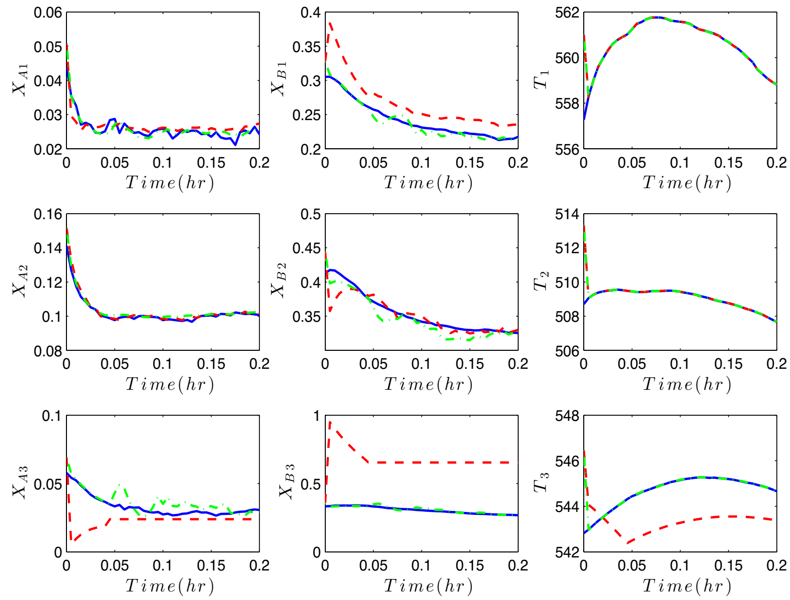

Figure 5.6: Communication Failure Case 1: State trajectories of the actual state (solid line), estimated states given by the proposed CDMHE (dashed line) and the centralized MHE (dash-dotted line).

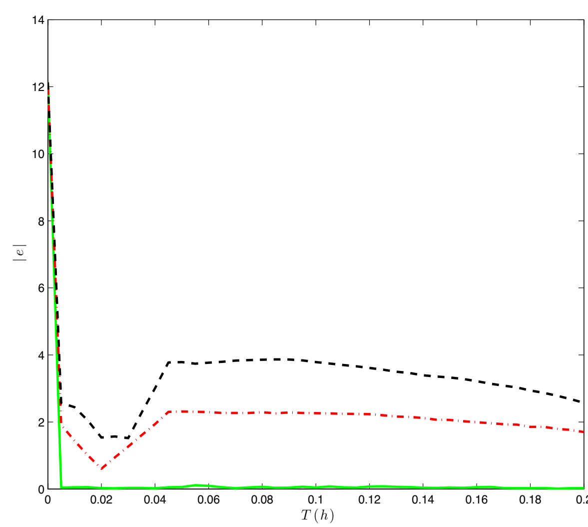

Figure 5.7: Communication Failure Case 1: Trajectories of the error norm given by proposed CDMHE (dash-dotted line), the centralized MHE (solid line), and the decentralized MHE (dashed line).

of the prediction-driven CDMHE is still much better than the performance of the decentralized MHE which can be seen from Figure 5.8 and Figure 5.10.

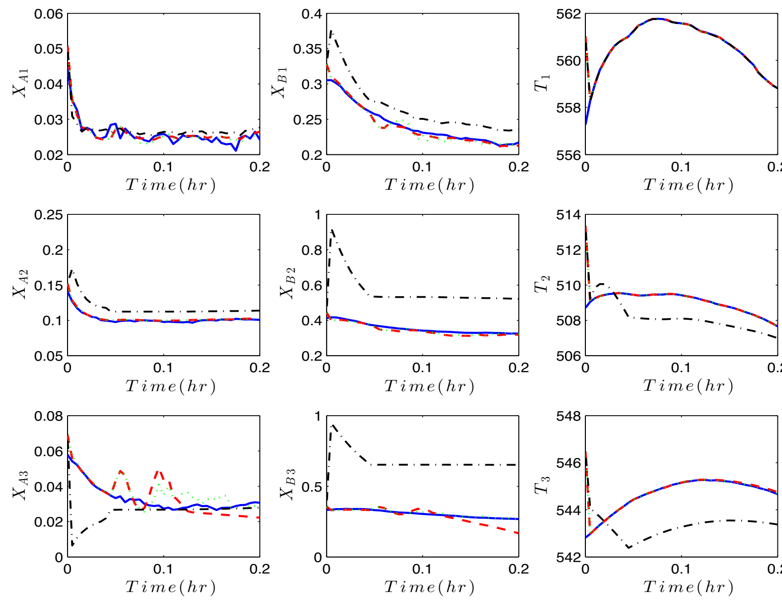

Figure 5.8: Communication Failure Case 2: State trajectories of the actual state (solid line), estimated states given by the proposed CDMHE (dashed line), the centralized MHE (dotted line), and the decentralized MHE (dash-dotted line).

From the above simulations in the two case studies, the proposed prediction-driven CDMHE shows resiliency against communication failure. When the communication failure happens, the subsystem that loses connection with the coordinator works in a decentralized way while other subsystems that coupled with this subsystems are affected. Therefore, the estimated states given by the prediction-driven CDMHE cannot converge to the estimated states given by the centralized MHE.

5.3 Premature Termination of the Coordination Algorithm

Premature termination of the coordination algorithm means that during sampling time k, the coordination algorithm stops at iteration 's' before the estimated state given by the coordination algorithm converges to the estimated state given by the centralized MHE. The premature termination of the coordination algorithm may happen for various reasons. For example, the time interval is not sufficient long to reach

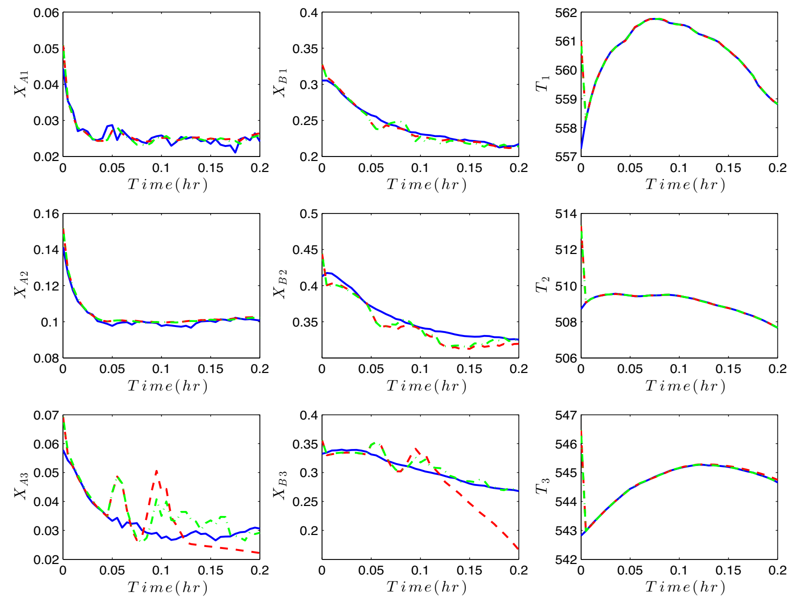

Figure 5.9: Communication Failure Case 2: State trajectories of the actual state (solid line), estimated states given by the proposed CDMHE (dashed line) and the centralized MHE (dash-dotted line).

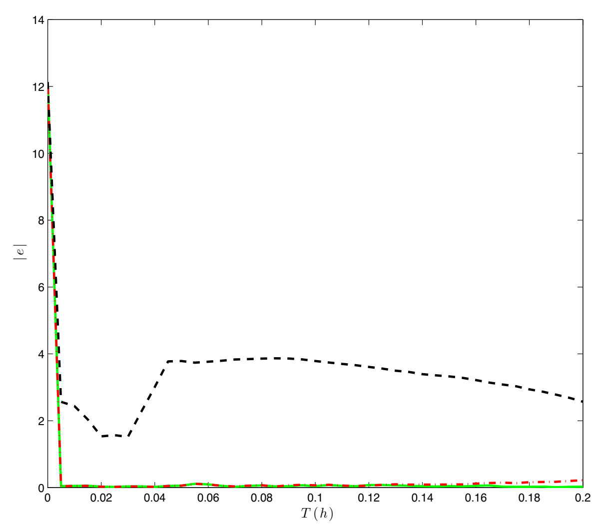

Figure 5.10: Communication Failure Case 2: Trajectories of the error norm given by proposed CDMHE (dash-dotted line), the centralized MHE (solid line), and the decentralized MHE (dashed line).

the convergence or the stop criterion threshold ϵ is poorly chosen. In this section, we consider that the iteration stops at 's_k' in time interval k. Under the assumption that the proposed prediction-driven CDMHE converges, the estimated state $\hat{X}^{(s)}(k)$ given by the coordination algorithm converges to $\hat{X}^{*}(k)$ as iteration 's' increases. Thus premature termination at iteration 's_k' will cause $\hat{X}^{(s_k)}(k)$ belongs to the neighborhood of the centralized optimal solution.

5.3.1 Simulation Results

In this section, the chemical process with two-CSTR and one seperator introduced in Section 4.4.2 is used. From the simulation case in Section 4.4.2, it can be obtained from Figure 4.10 that the iteration numbers are between 8 to 22. So in this case, we fixed the iterations as 7, which means the coordinated algorithm is terminated before it reaches the termination threshold. The other simulation settings are the same as that in Section 4.4.2. The results are shown in Figure 5.11 - Figure 5.13. Figure 5.11 shows the estimated state trajectories given by the prediction-driven CDMHE with premature termination, the centralized MHE, the decentralized MHE and the actual states. From Figure 5.12, we can see that the estimated states given by the prediction-driven CDMHE with premature termination are different from the centralized ones, but they are very close especially after $0.05h$. The reason is that the actual iterations after $t = 0.05h$ are around 11 which is closer to 7. From Figure 5.11 and Figure 5.13, we can see that the estimated states given by the prediction-driven CDMHE with premature termination are still much better than the decentralized MHE.

5.4 Conclusions

In this chapter, three scenarios are studied based on the proposed prediction-driven CDMHE algorithm in Chapter 4. A triggered communication algorithm for prediction-driven CDMHE was presented in Section 5.1. The communications between subsystems and the coordinator were triggered by the difference between the current estimated state and the last sent estimated state. The proposed prediction-driven CDMHE with triggered communication reduced the communication cost between the subsystems and the coordinator. The algorithm was applied to a chemical process in

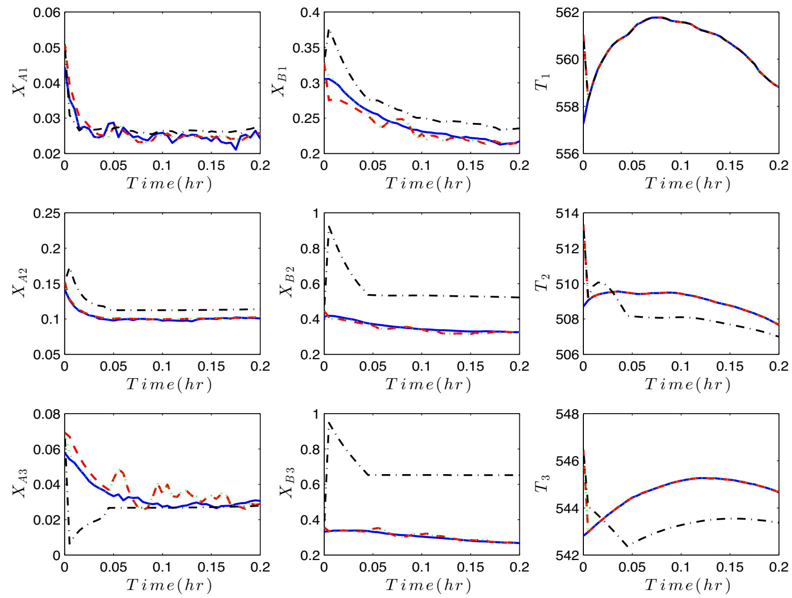

Figure 5.11: Premature termination case: State trajectories of the actual state (solid line), estimated states given by the proposed CDMHE (dashed line), the centralized MHE (dotted line), and the decentralized MHE (dash-dotted line).

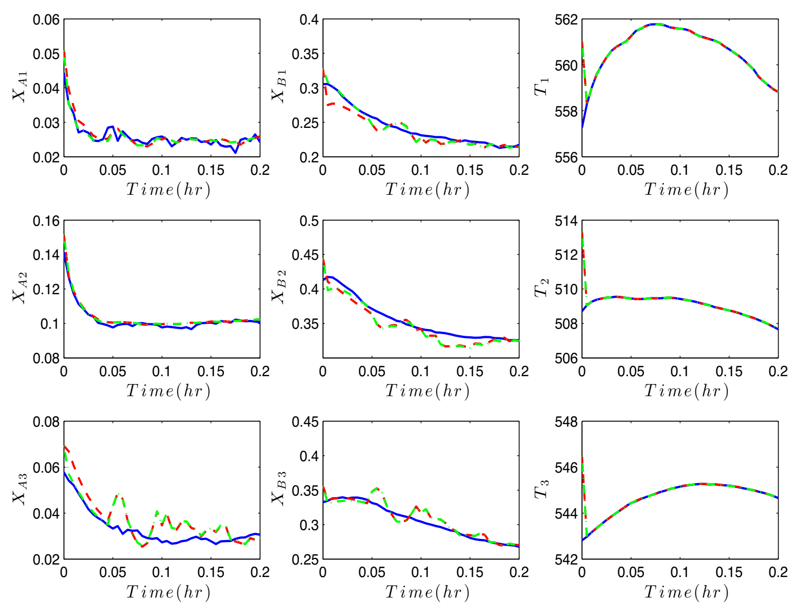

Figure 5.12: Premature termination case: State trajectories of the actual state (solid line), estimated states given by the proposed CDMHE (dashed line), and the centralized MHE (dash-dotted line).

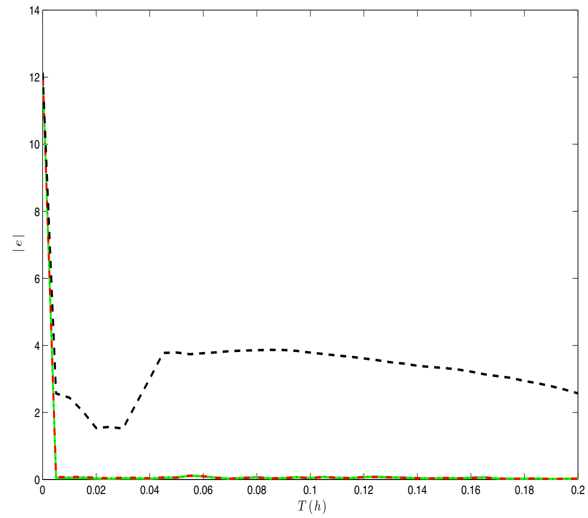

Figure 5.13: Premature termination case: Trajectories of the error norm given by proposed CDMHE (dash-dotted line), the centralized MHE (solid line), and the decentralized MHE (dashed line).

Section 5.1.2. The results were compared with the results with the prediction-driven CDMHE with regular communication from a communication cost point of view. With appropriate triggering thresholds, the proposed triggered communication algorithm not only keeps the estimation performance but also reduces the communication cost between subsystems and the coordinator. Subsequently, two cases of communication failure were discussed in Section 5.2. The prediction-driven CDMHE showed autonomy and resiliency against communication failures. In the first case, a subsystem lost connection with the coordinator before the communication began. In the other case, the subsystem lost connection with the coordinator after information been exchanged. The corresponding simulation results were given in Section 5.2.1. The results of both cases showed that the proposed prediction-driven CDMHE algorithm can handle the communication failure and provide better estimates than the decentralized MHE. Finally, premature termination of the coordination algorithm was studied in Section 5.3. The premature termination of coordination algorithm rendered the estimated states belongs to a neighborhood of the centralized optimal solution. The simulation results showed that the proposed prediction-driven CDMHE could keep the performance close to the centralized one.

Chapter 6

Conclusions

6.1 Summary

The focus of this thesis was on the development of the coordinated distribution moving horizon state estimation schemes for large-scale systems. Specially, the class of systems considered is a class of linear systems that composed of several subsystems that interacts with each other via their states. The CDMHE algorithms proposed in this work are intended to coordinate the local MHEs to achieve the optimal plant-wide performance, that is, the centralized state estimation performance, while maintain the flexibility of the decentralized estimation framework. Both the CDMHE schemes proposed in this thesis give guaranteed convergence to the centralized optimal solution when the convergence conditions are satisfied. The coordinated distributed estimators can be constructed with minor modifications to the exsiting decentralized estimators. The effectiveness of the coordinated distributed estimation schemes is shown in this thesis by implementing the algorithms into chemical processes.

In Chapter 3, a price-driven coordinated algorithm was developed for the distributed moving horizon state estimation, where the local MHE estimates the process states, noises and interactions. However, it was shown that the conventional price-driven coordinated algorithm cannot be used for state estimation purpose since it requires measurement of the full state vector. An improved price-driven CDMHE was proposed to address the issue of the standard version. In the improved price-driven coordinated algorithm, the local MHE recieves price vector from coordinator and sends the estimated state, noise, interaction and sensitivity matrix to the coordinator. Coordinator uses these information to calculate the price vector. Newton's method is used to update the price vector. Without consideration of the inequality constraints, the proposed price-driven CDMHE algorithm was shown to converge in two iterations. For the system with inequality constraints, the sensitivity matrix is divided into active set and inactive set, and changes with the active set. Due to the existence of the inequality constraints, more iterations are needed to achieve the convergence.

The drawback of the improved price-driven CDMHE is the additional term $\frac{1}{2} \| \hat{H}_i(k) \|_{\boldsymbol{D}_i^{-1}}$ in the cost function. This term is added purely to ensure the singular solution does not arise in the local MHEs. Therefore, to avoid this drawback, in Chapter 4, a prediction-driven coordinated algorithm was developed for the distributed moving horizon state estimation, where the local MHE estimates the state and process noise. In this coordinated scheme, the local MHE receives estimated interaction and price vector from the coordinator and sends estimated state and noise to the coordinator. The proposed prediction-driven algorithm was shown to converge to the centralized optimal solution with the increasing of the iteration number. A barrier function method was proposed to handle inequality constraints. In Chapter 5, the robustness of the prediction-driven CDMHE is further investigated under different scenarios, including triggered communication, communication failure and premature termination. The proposed prediction-driven CDMHE algorithm showed robustness and resilience to these cases.

It should be noted that both Chapter 3 and Chapter 4 focused on the development of the coordination algorithms of the local MHEs. Different coordination methods were used in these chapters. Without consideration of the inequality constraints, price-driven CDMHE algorithm shows fast convergence by using Newton's method to update the price vector, while the prediction-driven CDMHE algorithm may need more iterations to achieve the centralized MHE performance. When these two coordinated algorithms both converge, the corresponding price vector converges to the Lagrange multiplier that associate with the system equality constraint of the centralized MHE.

The key point of the coordination methods lies in the definition and coordination of the subproblems. In this thesis, the two proposed coordination methods, price-driven CDMHE and prediction-driven CDMHE both achieve the plant-wide optimal performance. There are still numerous challenges remain in the application of the coordinated distributed state estimation schemes that need to be addressed.

6.2 Directions for Future Work

Based on the work presented in the previous chapters, some possible directions for future work are listed as follows:

- Both coordination methods proposed in this thesis need information exchange between subsystems and coordinator, and the iterations during each sampling interval have great impact on the computation and communication load. Therefore, the complexity study for both price-driven CDMHE and prediction-driven CDMHE is recommended Knowing the computation load of different CDMHE methods can give us a better understanding of the coordination algorithms and help us to choose algorithm from the candidates;

- Although the two CDMHE algorithms both converge to the centralized optimal performance, arrival cost is not considered in both cases. Arrival cost summarizes the effect of previous data. A study that includes arrival cost in the cost function is needed;

- In order to converge to the centralized MHE optimal solution, the prediction-driven CDMHE needs to satisfy the convergence condition proposed in Chapter 4. The convergence condition is related to the system matrices, weighting matrices and size of the horizon. The system matrices represent the interaction strength in some way. The effects of the interaction strength, weighting matrices and size of the horizon on the convergence condition are worth investigating;

- The focus of this thesis is linear systems. The work can be extended to nonlinear systems by using the same coordination methods.

- The integration of CDMHE with distributed predictive control (e.g., DMPC) needs investigation. DMPC has been widely applied to the control of large-scale complex system, which requires the state measurement every sampling time.

The CDMHE algorithms can provide state estimates to the control network when the state measurements are not available.

Bibliography

Al-Gherwi, W., H. Budman and A. Elkamel (2013). A robust distributed model predictive control based on a dual-mode approach. *Computers & Chemical Engineering* **50**, 130–138.

Bakule, L. (2008). Decentralized control: An overview. *Annual Reviews in Control* **32**, 87–98.

Boyd, S. and L. Vandenberghe (2004). *Convex optimization.* Cambridge university press.

Cheng, R. (2007). Decomposition and coordination of large-scale operations optimization. PhD thesis. University of Alberta.

Cheng, R., J. F. Forbes and W. S. Yip (2007). Price-driven coordination method for solving plant-wide MPC problems. *Journal of Process Control* **17**, 429–438.

Chong, E.K. and S.H. Zak (2013). *An introduction to optimization.* Vol. 76. John Wiley & Sons.

Christofides, P. D., R. Scattolini, D. Muñoz de la Peña and J. Liu (2013). Distributed model predictive control: A tutorial review and future research directions. *Computers & Chemical Engineering* **51**, 21–41.

Cohen, G. (1977). On an algorithm of decentralized optimal control.. *Journal of Mathematical Analysis and Applications* **59**, 242–259.

Farina, M., G. Ferrari-Trecate and R. Scattolini (2010). Distributed moving horizon estimation for linear constrained systems. *IEEE Transactions on Automatic Control* **55**, 2462–2475.

Farina, M., G. Ferrari-Trecate and R. Scattolini (2012). Distributed moving horizon estimation for nonlinear constrained systems. *International Journal of Robust and Nonlinear Control* **22**, 123–143.

Khan, U. A. and J. M. F. Moura (2008). Distributing the Kalman filter for large-scale systems. *IEEE Transactions on Signal Processing* **56**, 4919 – 4935.

Liu, J., D. Muñoz de la Peña and P. D. Christofides (2009). Distributed model predictive control of nonlinear process systems. *AIChE Journal* **55**, 1171–1184.

Liu, J., Ohran B.J. Christofides P.D. de la Peña, D.M and J. F. Davis (2008). A two-tier architecture for networked process control. *Chemical Engineering Science* **63**(22), 5394–5409.

Marcos., N. (2011). Coordinated-Distributed Optimal Control of Large-Scale Linear Dynamic Systems. PhD thesis. University of Alberta.

Mirsky, L. (2012). *An introduction to linear algebra*. Courier Corporation.

Mohseni., P. G. (2013). Coordination Techniques for Distributed Model Predictive Control. PhD thesis. University of Alberta.

Mutambara, A. G. O. and H. E. Durrant-Whyte (2000). Estimation and control for a modular wheeled mobile robot. *IEEE Transactions on Control Systems Technology* **8**, 35–46.

Scattolini, R. (2009). Architectures for distributed and hierarchical model predictive control - A review. *Journal of Process Control* **19**, 723–731.

Silvester, J.R. (2000). Determinants of block matrices. *The Mathematical Gazette* **84**(501), 460–467.

Stanković, S. S., M. S. Stanković and D. M. Stipanović (2009). Consensus based overlapping decentralized estimator. *IEEE Transactions on Automatic Control* **54**, 410–415.

Stewart, B. T., S. J. Wright and J. B. Rawlings (2011). Cooperative distributed model predictive control for nonlinear systems. *Journal of Process Control* **21**, 698–704.

Sun, Y. and N. H. El-Farra (2008). Quasi-decentralized model-based networked control of process systems. *Computers and Chemical Engineering* **32**, 2016–2029.

Tippett, M. J. and J. Bao (2013). Distributed model predictive control based on dissipativity. *AIChE Journal* **59**, 787–804.

Zhang, F. (2006). *The Schur complement and its applications*. Vol. 4. Springer Science & Business Media.

Zhang, J. and J. Liu (2013). Distributed moving horizon estimation for nonlinear systems with bounded uncertainties. *Journal of Process Control* **23**, 1281–1295.

Appendix A

Invertibility of Matrices

Prior to the proof, the term **Schur complement** and some related lemmas are introduced. Assume that a square matrix $\boldsymbol{M} \in \mathbb{R}^{(l+m)\times(l+m)}$is partitioned into the following four blocks:

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{T} \\ \boldsymbol{U} & \boldsymbol{V} \end{bmatrix} \tag{A.1}$$

where $\boldsymbol{S} \in \mathbb{R}^{l\times l}$, $\boldsymbol{T} \in \mathbb{R}^{l\times m}$, $\boldsymbol{U} \in \mathbb{R}^{m\times l}$ and $\boldsymbol{V} \in \mathbb{R}^{m\times m}$. If the square matrix $\boldsymbol{S}$ is nonsingular, then:

$$\boldsymbol{M}/\boldsymbol{S} \triangleq \boldsymbol{V} - \boldsymbol{U}\boldsymbol{S}^{-1}\boldsymbol{T} \tag{A.2}$$

is defined to be the **Schur complement** of $\boldsymbol{M}$ relative in $\boldsymbol{S}$. Similarly, if the square matrix $\boldsymbol{V}$ is nonsingular, the **Schur complement** of $\boldsymbol{M}$ relative in $\boldsymbol{V}$ is defined as:

$$\boldsymbol{M}/\boldsymbol{V} \triangleq \boldsymbol{S} - \boldsymbol{T}\boldsymbol{V}^{-1}\boldsymbol{U} \tag{A.3}$$

The **Schur complement** is a very useful tool in matrix analysis. A commonly used property of **Schur complement** is presented in **Lemma** A.0.1:

Lemma A.0.1. *(Zhang, 2006) Let $\boldsymbol{M} \in \mathbb{R}^{(l+m)\times(l+m)}$ be a square matrix partitioned as in equation A.1. If $\boldsymbol{S}$ is nonsingular, then:*

$$det(\boldsymbol{M}) = det(\boldsymbol{S})det(\boldsymbol{M}/\boldsymbol{S}) \tag{A.4}$$

Similarly, if $\mathbf{V}$ is nonsingular, there is:

$$det(\boldsymbol{M}) = det(\boldsymbol{V})det(\boldsymbol{M}/\boldsymbol{V}) \tag{A.5}$$

where $det(\boldsymbol{M})$ means the determinant of matrix $\boldsymbol{M}$.

When $\boldsymbol{S}(\boldsymbol{V})$ is nonsingular, that is $det(\boldsymbol{S}) \neq 0$ ($det(\boldsymbol{V}) \neq 0$). Then in equation (A.4) (equation (A.5)), $det(\boldsymbol{M}) \neq 0$ is equivalent to $det(\boldsymbol{M}/\boldsymbol{S}) \neq 0$ ($det(\boldsymbol{M}/\boldsymbol{V}) \neq 0$), which leads to the following Corollary:

Corollary A.0.1. *For the partitioned matrix $\boldsymbol{M}$ as in (A.1), if the block $\boldsymbol{S}(\boldsymbol{V})$ is nonsingular, then $\boldsymbol{M}$ is invertible if and only if the* ***Schur complement*** *$\boldsymbol{M}/\boldsymbol{S}(\boldsymbol{M}/\boldsymbol{V})$ is invertible.*

A.1 The Invertibility Discussion of $\mathit{\Lambda}_i$

We use Schur complement to prove the invertibility of $\mathit{\Lambda}_i$.

$$\Lambda_i = \left[\begin{array}{cc:cc} C_{ii}^T R_{ii}^{-1} C_{ii} & 0 & 0 & G_{A_{ii}}^T \\ 0 & Q_{ii}^{-1} & 0 & -I \\ \hdashline 0 & 0 & 0 & -I \\ G_{A_{ii}} & -I & -I & 0 \end{array}\right]$$

So we can write the matrix as:

$$S = \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix},\ T = \begin{bmatrix} 0 & G_{A_{ii}}^T \\ 0 & -I \end{bmatrix},\ U = \begin{bmatrix} 0 & 0 \\ G_{A_{ii}} & -I \end{bmatrix},\ V = \begin{bmatrix} 0 & -I \\ -I & 0 \end{bmatrix}$$

$$\begin{aligned} M/V &= S - TV^{-1}U \\ &= \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix} - \begin{bmatrix} 0 & G_{A_{ii}}^T \\ 0 & -I \end{bmatrix} \begin{bmatrix} 0 & -I \\ -I & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ G_{A_{ii}} & -I \end{bmatrix} \\ &= \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix} - \begin{bmatrix} -G_{A_{ii}}^T & 0 \\ I & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ G_{A_{ii}} & -I \end{bmatrix} \\ &= \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix} \end{aligned}$$

As $det(M) = det(V)det(M/V)$, we need to prove $det(M/V) \neq 0$.

$$det(M/V) = det\{\begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix}\} = det\{C_{ii}^T R_i^{-1} C_{ii}\} det\{C_{ii}^T\}$$

We need to prove $det\{C_{ii}^T R_i^{-1} C_{ii}\} \neq 0$.

$$det\{C_{ii}^T R_i^{-1} C_{ii}\} = det\{C_{ii}^T R_i^{-1/2}\} det\{R_i^{-1/2} C_{ii}\}$$

$$\begin{aligned} det\{C_{ii}^T R_i^{-1/2}\} &= det\{R_i^{-1/2} C_{ii}\} \\ &= det\{C_{ii}\} det\{R_i\}^{-1/2} \end{aligned}$$

We need to prove $det\{C_{ii}\} \neq 0$. However, only when C_{ii} is full rank, $det\{C_{ii}\} \neq 0$. Thus $\mathit{\Lambda}_i$ is invertible only when C_{ii} is full rank.

A.2 Invertibility Condition of Improved Price-driven CDMHE

After adding term $\frac{1}{2}\hat{H}_i(k)^T D_i^{-1}\hat{H}_i(k)$ to the cost function in the price-driven coordinated local MHE design, Λ_i defined in equation (3.30) becomes:

$$\Lambda_i = \begin{bmatrix} \Xi_i^* & F_i^T \\ F_i & 0 \end{bmatrix} \tag{A.6}$$

Substituting the Ξ_i^* and F_i^T into Λ_i, the following equation can be obtained:

$$\Lambda_i = \left[\begin{array}{cc:cc} C_{ii}^T R_i^{-1} C_{ii} & 0 & 0 & G_{A_{ii}}^T \\ 0 & Q_i^{-1} & 0 & -I \\ \hdashline 0 & 0 & D_i^{-1} & -I \\ G_{A_{ii}} & -I & -I & 0 \end{array}\right] \tag{A.7}$$

Using Schur complement to discuss the invertibility condition that D_i^{-1} needs to satisfy, we can write the matrix as:

$$S = \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix},\ T = \begin{bmatrix} 0 & G_{A_{ii}}^T \\ 0 & -I \end{bmatrix},\ U = \begin{bmatrix} 0 & 0 \\ G_{A_{ii}} & -I \end{bmatrix},\ V = \begin{bmatrix} D_i^{-1} & -I \\ -I & 0 \end{bmatrix} \tag{A.8}$$

Therefore the inverse of the matrix V is:

$$V^{-1} = \begin{bmatrix} 0 & -I \\ -I & -D_i^{-1} \end{bmatrix} \tag{A.9}$$

Therefore, the following expression can be obtained:

$$\begin{aligned} M/V &= S - TV^{-1}U \\ &= \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix} - \begin{bmatrix} 0 & G_{A_{ii}}^T \\ 0 & -I \end{bmatrix} \begin{bmatrix} 0 & -I \\ -I & -D_i^{-1} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ G_{A_{ii}} & -I \end{bmatrix} \\ &= \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} & 0 \\ 0 & Q_i^{-1} \end{bmatrix} - \begin{bmatrix} -G_{A_{ii}}^T D_i^{-1} G_{A_{ii}} & G_{A_{ii}}^T D_i^{-1} \\ D_i^{-1} G_{A_{ii}} & -D_i^{-1} \end{bmatrix} \\ &= \begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} + G_{A_{ii}}^T D_i^{-1} G_{A_{ii}} & G_{A_{ii}}^T D_i^{-1} \\ D_i^{-1} G_{A_{ii}} & Q_i^{-1} + D_i^{-1} \end{bmatrix} \end{aligned} \tag{A.10}$$

As $det(M) = det(V)det(M/V)$ and V is invertible ($det(V) \neq 0$), therefore we need to find the condition that $det(M/V) \neq 0$. By using Schur complement again, we can get that:

$$\begin{aligned} det(M/V) &= det\{\begin{bmatrix} C_{ii}^T R_i^{-1} C_{ii} + G_{A_{ii}}^T D_i^{-1} G_{A_{ii}} & G_{A_{ii}}^T D_i^{-1} \\ D_i^{-1} G_{A_{ii}} & Q_i^{-1} + D_i^{-1} \end{bmatrix}\} \\ &= det(Q_i^{-1} + D_i^{-1}) det\{C_{ii}^T R_i^{-1} C_{ii} + G_{A_{ii}}^T D_i^{-1} G_{A_{ii}} - G_{A_{ii}}^T D_i^{-1} (Q_i^{-1} + D_i^{-1})^{-1} D_i^{-1} G_{A_{ii}}\} \\ &= det(Q_i^{-1} + D_i^{-1}) det\{C_{ii}^T R_i^{-1} C_{ii} + G_{A_{ii}}^T (D_i^{-1} - D_i^{-1}(Q_i^{-1} + D_i^{-1})^{-1} D_i^{-1}) G_{A_{ii}}\} \end{aligned} \tag{A.11}$$

Since Q_i^{-1} and D_i^{-1} are the weighting matrices which are usually picked as symmetric positive definite matrices, the determinant of $(Q_i^{-1} + D_i^{-1})$ cannot be zero. In order to make the matrix Λ_i in (A.6) invertible, the weighting matrix D_i^{-1} needs to make the determinant of $\{C_{ii}^T R_i^{-1} C_{ii} + G_{A_{ii}}^T (D_i^{-1} - D_i^{-1}(Q_i^{-1} + D_i^{-1})^{-1} D_i^{-1}) G_{A_{ii}}\}$ not equal to zero.

Therefore when we pick the weighting matrix D_i^{-1}, the following condition should be satisfied:

$$det\{C_{ii}^T R_i^{-1} C_{ii} + G_{A_{ii}}^T (D_i^{-1} - D_i^{-1}(Q_i^{-1} + D_i^{-1})^{-1} D_i^{-1}) G_{A_{ii}}\} \neq 0 \tag{A.12}$$

A.3 Uniqueness Solution of the Prediction-driven CDMHE

For the unconstrained prediction-driven CDMHE, in the s^{th} communication cycle, the Lagrange equation for subsystem i can be expressed as following:

$$\begin{aligned}\mathcal{L}_i =& \frac{1}{2} \begin{bmatrix}\hat{X}_i(k)^T & \hat{W}_i(k)^T\end{bmatrix} \begin{bmatrix}\boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} & 0 \\ 0 & \boldsymbol{Q}_i^{-1}\end{bmatrix} \begin{bmatrix}\hat{X}_i(k) \\ \hat{W}_i(k)\end{bmatrix} + \{\begin{bmatrix}-Y_i^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} & 0\end{bmatrix} \\ &+ \boldsymbol{p}^{(s)^T} \Theta_i\} \begin{bmatrix}\hat{X}_i(k) \\ \hat{W}_i(k)\end{bmatrix} + \lambda_i^T (G_i^{eq} \begin{bmatrix}\hat{X}_i(k) \\ \hat{W}_i(k)\end{bmatrix} - \hat{H}_i(k))\end{aligned} \tag{A.13}$$

The optimal solution of problem (4.2) at s^{th} communication cycle is:

$$\begin{cases} \dfrac{\partial \mathcal{L}_i}{\partial \hat{X}_i(k)} = \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} \hat{X}_i(k) - \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} Y_i + \Theta_{A_i}^T \boldsymbol{p}^{(s)} + G_{A_{ii}}^T \lambda_i = \emptyset \\ \dfrac{\partial \mathcal{L}_i}{\partial \hat{W}_i(k)} = \boldsymbol{Q}_i^{-1} \hat{W}_i(k) + G_{B_{ii}}^T \lambda_i = \emptyset \\ \dfrac{\partial \mathcal{L}_i}{\partial \hat{\lambda}_i(k)} = G_{A_{ii}} \hat{X}_i(k) + G_{B_{ii}} \hat{W}_i(k) = -\sum_{j \neq i} G_{A_{ij}} X_j^{(s)}(k) + G_{B_{ij}} W_j^{(s)}(k) \end{cases} \tag{A.14}$$

where Θ_i can be divided into two parts, that is $\Theta_i = [\Theta_{A_i}, \Theta_{B_i}]$, $G_i^{eq} = [G_{A_{ii}}^{eq}, G_{B_{ii}}^{eq}]$. $\Theta_{A_i}, \Theta_{B_i}$ are defined as:

To ensure the local MHE has unique solution, the matrix $\boldsymbol{M}_i$ must be invertible:

$$\boldsymbol{M}_i = \begin{bmatrix} \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} & 0 & G_{A_{ii}}^T \\ 0 & \boldsymbol{Q}_i^{-1} & G_{B_{ii}}^T \\ G_{A_{ii}} & G_{B_{ii}} & 0 \end{bmatrix} \tag{A.15}$$

We will use Schur complement to prove the invertibility of M_i. From the definition of $G_{B_{ii}}$, the matrix M_i can be expressed as:

$$\boldsymbol{M}_i = \left[\begin{array}{c:cc} \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} & 0 & G_{A_{ii}}^T \\ \hdashline 0 & \boldsymbol{Q}_i^{-1} & -I \\ G_{A_{ii}} & -I & 0 \end{array}\right] \tag{A.16}$$

So we can write the matrix as:

$$S = \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii},\ T = \begin{bmatrix} 0 & G_{A_{ii}}^T \end{bmatrix},\ U = \begin{bmatrix} 0 \\ G_{A_{ii}} \end{bmatrix},\ V = \begin{bmatrix} \boldsymbol{Q}_i^{-1} & -I \\ -I & 0 \end{bmatrix}.$$

Therefore, the following equation can be obtained:

$$\begin{aligned} M/V &= S - TV^{-1}U \\ &= \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} - \begin{bmatrix} 0 & G_{A_{ii}}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_i^{-1} & -I \\ -I & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ G_{A_{ii}} \end{bmatrix} \\ &= \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} - \begin{bmatrix} 0 & G_{A_{ii}}^T \end{bmatrix} \begin{bmatrix} 0 & -I \\ -I & -\boldsymbol{Q}_i^{-1} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ G_{A_{ii}} \end{bmatrix} \\ &= \boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii} + G_{A_{ii}}^T \boldsymbol{Q}_i^{-1} G_{A_{ii}} \end{aligned} \tag{A.17}$$

Neither $\boldsymbol{C}_{ii}^T \boldsymbol{R}_i^{-1} \boldsymbol{C}_{ii}$ nor $G_{A_{ii}}^T \boldsymbol{Q}_i^{-1} G_{A_{ii}}$ is full rank; however, their summation may be full rank. From Section 2.5 in Chapter 2, it is shown that for an observable system consists of n states, if the horizon $H_p \geq n$, the MHE formulation used in Section 2.3 and Section 2.4 has a unique solution. The matrix M_i is the same with the one used in decentralized MHE, thus the invertibility of M_i can be guaranteed.

Appendix B

Coordinated Term Verification

Our purpose is to compute $\mathit{\Gamma}_i(\hat{X}^{(s)}(k))\hat{X}_i(k)$, or equivalently to proof $p^{(s)T}\Theta_i \begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} = \mathit{\Gamma}_i(\hat{X}^{(s)}(k))\hat{X}_i(k)$. We have seen that

$$\mathit{\Gamma}_i(\hat{X}^{(s)}(k)) = \frac{\partial \mathcal{J}_D}{\partial \hat{H}(k)}|_{(\hat{X}^{(s)}(k),K(\hat{X}^{(s)}(k)))}\frac{d\hat{H}(k)}{d\hat{X}_i(k)}|_{\hat{X}^{(s)}(k)} \tag{B.1}$$

From (4.35), we can express $\frac{\partial \mathcal{J}_D}{\partial \hat{H}(k)}|_{(\hat{X}^{(s)}(k),\hat{H}^{(s)}(k))}$ as $\frac{\partial \mathcal{J}_D}{\partial \hat{W}(k)}\frac{d\hat{W}(k)}{d\hat{H}(k)}|_{(\hat{X}^{(s)}(k),\hat{H}^{(s)}(k))}$. According to (4.35) and (4.2b), we can get that

$$\frac{\partial \mathcal{J}_D}{\partial \hat{W}(k)}\frac{d\hat{W}(k)}{d\hat{H}(k)}|_{(\hat{X}^{(s)}(k),\hat{H}^{(s)}(k))} = (Q^{-1}\hat{W}^{(s)})^T\bar{G}_B^{-1} \tag{B.2}$$

Since $G_{ij} = [G_{A_{ij}}, G_{B_{ij}}]$, $G_{B_{ij}} = 0$, from (4.3), we can express $\Theta_i = [\Theta_{iA}, \Theta_{iB}]$, while Θ_{iB} consists of $G_{B_{ij}}$, so that $\Theta_{iB} = 0$. Then $\hat{H}^{(s)}(k) = [\hat{H}_1^{(s)}(k)^T, \hat{H}_2^{(s)}(k)^T, ..., \hat{H}_N^{(s)}(k)^T]^T$, $\hat{H}^{(s)}(k)$ can be expressed as following:

$$\begin{aligned}\hat{H}^{(s)}(k) &= -\begin{bmatrix}\Theta_{1A}, & \Theta_{2A}, & \cdots, & \Theta_{NA}, & \Theta_{1B}, & \Theta_{2B}, & \cdots, & \Theta_{NB}\end{bmatrix}\begin{bmatrix}\hat{X}^{(s)}(k)\\ \hat{W}^{(s)}(k)\end{bmatrix}\\ &= -\begin{bmatrix}\bar{\Theta}_A & \bar{\Theta}_B\end{bmatrix}\begin{bmatrix}\hat{X}^{(s)}(k)\\ \hat{W}^{(s)}(k)\end{bmatrix} = -\bar{\Theta}_A\hat{X}^{(s)}(k)\end{aligned} \tag{B.3}$$

From the expression of $\hat{H}^{(s)}(k)$, we can get the expression of $\frac{d\hat{H}(k)}{d\hat{X}_i(k)}|_{\hat{X}^{(s)}(k)}$ as:

$$\frac{d\hat{H}(k)}{d\hat{X}_i(k)}|_{\hat{X}^{(s)}(k)} = -\Theta_{iA} \tag{B.4}$$

Thus we get $\mathit{\Gamma}_i(\hat{X}^{(s)}(k))\hat{X}_i(k)$ as following:

$$\mathit{\Gamma}_i(\hat{X}^{(s)}(k))\hat{X}_i(k) = -(Q^{-1}\hat{W}^{(s)})^T\bar{G}_B^{-1}\Theta_{iA}\hat{X}_i(k) \tag{B.5}$$

From (4.9), we can get the expression of $p^{(s)^T}\Theta_i \begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix}$ as:

$$p^{(s)^T}\Theta_i \begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} = -(Q^{-1}\hat{W}^{(s)})^T G_B^{-1}\Theta_{iA}\hat{X}_i(k) \tag{B.6}$$

From the definition of $G_B,\ \bar{G}_B,\ \bar{\Theta}_B$, we can easily get that $G_B = \bar{G}_B + \bar{\Theta}_B$, since $\bar{\Theta}_B = 0$, we can get $G_B = \bar{G}_B$. So we can get the conclusion that:

$$p^{(s)^T}\Theta_i \begin{bmatrix}\hat{X}_i(k)\\ \hat{W}_i(k)\end{bmatrix} = \mathit{\Gamma}_i(\hat{X}^{(s)}(k))\hat{X}_i(k) \tag{B.7}$$