Data-based Techniques to Improve State Estimation in Model Predictive Control

by

Murali Rajaa Rajamani

A dissertation submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

(Chemical Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON

2007

To my family:

Dearest Dad, Mom, Brother and Paati

for their endless love and unconditional support

Data-based Techniques to Improve State Estimation in Model Predictive Control

Murali Rajaa Rajamani

Under the supervision of Professor James B. Rawlings

At the University of Wisconsin–Madison

Specifying the state estimator in model predictive control (MPC) requires separate knowledge of the disturbances entering the states and the measurements, which is usually lacking. In this dissertation, we develop the Autocovariance Least-Squares (ALS) technique which uses the correlations between routine operating data to form a least-squares problem to estimate the covariances for the disturbances. The ALS technique guarantees positive semidefinite covariance estimates by solving a semidefinite programming (SDP) problem. Many efficient algorithms for solving SDPs are available in the literature. New and simple necessary and sufficient conditions for the uniqueness of the covariance estimates are presented. We also formulate the optimal weighting to be used in the least-squares objective in the ALS technique to ensure minimum variance in the estimates. A modification to the above technique is then presented to estimate the

stochastic disturbance structure and the minimum number of disturbances required to represent the data.

Simplifications to the ALS technique are presented to facilitate implementation. It is also shown that the choice of the disturbance model in MPC does not affect the closed-loop performance if appropriate covariances are used in specifying the state estimator. The ALS technique is used to estimate the appropriate covariances regardless of the plant's true unknown disturbance source and the resulting estimator gain is shown to compensate for an incorrect choice of the source of the disturbance.

The parallels between the ALS technique and the maximum likelihood estimation (MLE) technique are shown by formulating the MLE as an equivalent ALS optimization with a particular choice for the weighting in the ALS objective.

An industrial application of the ALS technique on a nonlinear blending drum model and industrial operating data is described and the results are shown to give improved state estimates as compared to rough industrial estimates for the covariances.

Moving horizon estimator (MHE) and particle filters (PF) are two common state estimators used with nonlinear models. We present a novel hybrid of the MHE with the PF to combine the advantages of both techniques. The improved performance of the hybrid MHE/PF technique is illustrated on an example.

Acknowledgments

I thank the Divine for giving me everything I have so far and for having blessed me with good fortune. I am indebted to a great many people for making my years in Madison and graduate school a priceless experience.

I begin with the following Sanskrit saying by Adi Shankara: *" Guru Brahma, Guru Vishnu, Guru Devo Maheshwara. Guru Sakshath Parambrahma, Tasmai Shri Gurave Namaha,"* translated in English as *"Guru is the creator Brahma, Guru is the preserver Vishnu, Guru is the destroyer Shiva. Guru is directly the supreme spirit – I offer my salutations to this Guru."* I consider my advisor Prof. Jim Rawlings to be my *Guru*, without whose guidance I would not have gained even the fraction of my achievements in Madison. I am amazed time and again by his clarity of thinking, depth of analysis and simplification to the fundamentals. I will miss my intellectual interactions with him and I hope I have another opportunity to work with him in the future. He has been an idol to me and I will strive to think and work as efficiently and quickly as he.

I thank Profs. Graham, Maravelias, Barmish and Wright for taking time to be on my thesis committee.

I reserve my deepest gratitude and love for my father. It was my Dad's wish that

I pursue my graduate studies in the USA, and is the only reason for my present stature. His efforts and support for our family will remain unparalleled and I often look to the heavens to seek his blessings. My mom has constantly showered me with affection and reduced the physical distance between us through her love, making it possible for me to stay away from home for so long. My brother, Subash Rajaa, has been a constant source of encouragement and advice throughout my graduate school. Without his comfort, love and words of wisdom my work here would have been impossible.

Mary Diaz has been the 'mother' to the Rawlings' group and I have been especially close to her. I will miss Mary and her goodies. John Eaton has been extremely helpful and patient in answering all my dumb Linux questions. Within the Rawlings' group, I have developed friendships and acquaintances with past and present members. I treasure my friendship with Paul Larsen who has been my office-mate for the past five years. He has endured me, answered my questions, listened to my opinions and offered relief to my rants at all times. Ethan Mastny has been a great and fun person to work with and has been the go-to person with answers to all questions. I have enjoyed interacting with Aswin Venkat within the group and outside at other times. I hope he remains the 'cool' dude that he always was. I thank Brian Odelson for helping me with getting acquainted with the research in the group and Eric Haseltine for being the senior with helpful advice and suggestions at all times. In the relatively short time I had with the new members, Brett Stewart and Rishi Amrit, I believe the group has a bright future. I wish them the best for their remaining years in the group. Rishi, being the youngest member in the

group, has been helpful with simulations and particularly sporting with my practical jokes on him. Among other past and present graduate students in the department, I have found conversations with Aaron Lowe, Sebastian Hensel, Cliff Amundsen, Paul Bandi and Alexander Lutz have been a good relief during days of stressful work.

Over the five years in the Madison, I have made many life-long friends, who made my times in Madison extremely enjoyable. I will miss the unforgettable times with Manan Chopra and his opinions on every imaginable topic. I hope to see him in Wall Street soon and to continue my friendship with him in the future. I have been fortunate to have met Pallavi Umarji and I thank her for her company and friendship over the last couple of years. Soujanya Jampala has been a great friend in Madison and I wish the best for her life and career. Anand Nilekar, Rahul Nabar, Somen Saha and Ronak Vakil have been dependable friends and have providing me with timely advice and help. Santosh Reddy and Manish Rathi have always been welcoming for a night of poker and relief at their popular 'farmhouse'. Over the last year or so Pratik Pranay, Rohit Malshe, Shailesh Moolya, Uday Bapu, Vikram Adhikarla and Arvind Manian have given good company. Finally, Sury, my roommate for four years, deserves a special mention for being a brother-like figure to me in Madison.

MURALI RAJAA RAJAMANI

University of Wisconsin–Madison

October 2007

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A model of a system is one that accurately represents its dynamic behavior. With a variety of models available, the preferred choice is often the one that provides maximum flexibility and ease of use. Once a model is chosen, the next step is to choose a method to estimate the parameters for the model. Parameter estimation for models is a well researched topic, although the best model and the best estimation method are still debated and the choice is often governed by the practicality of a method (Ljung, 1987). The choice of the model in this dissertation work is the state-space model because of its flexibility in representing a wide variety of special cases, which require separate consideration in other models. For example, the ARMA models are applicable only for stationary systems. Nonstationary systems require an additional differencing term giving rise to ARIMA models (Box and Jenkins, 1976). State-space models are applicable for both stationary and nonstationary systems. In addition, periodicity and time delays are handled conveniently in state-space models by augmentation of the states. For

example, the linear state-space discrete-time model is given as:

$$x_{k+1} = Ax_k + Bu_k + Gw_k$$

$$y_k = Cx_k + v_k$$

where, the subscript 'k' denotes sampling of the values at times $t_k = k\Delta t$ with Δt being the sampling time. $x_k \in \mathbb{R}^{n\times 1}$ is a physical state of the system that is usually modelled from first principles. $y_k \in \mathbb{R}^{p\times 1}$ is the vector of the measured states of the systems. Both the states and the measurements are corrupted by additive noises w_k and v_k respectively. The inputs to the system are $u_k \in \mathbb{R}^{m\times 1}$. Such a discrete state-space model can be easily developed from first principles ordinary differential equations and discretized with the sampling time of Δt.

Model Predictive Control (MPC) is one of the most widely used multivariable control technique in the chemical industry. The use of the state-space model formulation in MPC is convenient in systems having multiple inputs and multiple outputs, as often is the case in even the simplest of chemical units. MPC consists of three components: the regulator, the state estimator and the target calculator. A schematic diagram of the different units making up MPC is shown in Figure 1.1.

The state estimator unit in MPC provides feedback from past measurements and inputs to optimally estimate the state and disturbances. Usually, the deterministic system matrices (A, B, C) are estimated as a part of an identification scheme. The optimality of the state estimator is then based on an accurate knowledge of the actual noise statistics entering the plant. For illustrative purposes, if we assume the noises w_k and

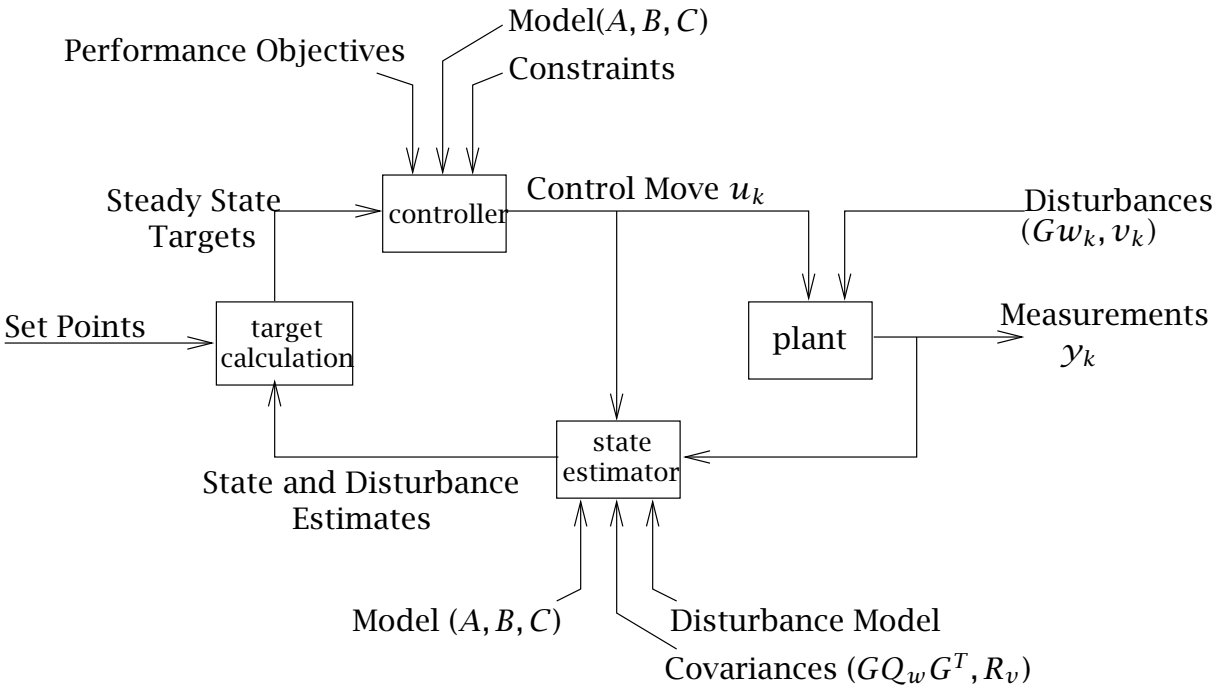


Figure 1.1: Schematic diagram with the different units involved in the Model Predictive Control (MPC) scheme

v_k to be drawn from the Gaussian distribution with zero means and covariances Q_w and R_v, as seen in Figure 1.1, the state estimator requires accurate information about GQ_wG^T, R_v in order to give an optimal state estimate. Usually this knowledge is not available and needs to be estimated from closed-loop measurements and inputs from the operating plants. Current industrial practice is to set these disturbance/noise covariances heuristically, which then results in suboptimal state estimates. The objective of this dissertation work is to develop a technique that would estimate these disturbance/noise covariances from closed-loop operating data. Once we have this knowledge, the state estimator can then be specified optimally with the estimated covariances

leading to better closed-loop control. Finding the number of independent disturbances entering the plant is another issue that is addressed in this work. If the number of independent disturbances is not accurate, the result is overmodelling and nonuniqueness in the covariance estimation. Monitoring and state estimation tools work best using models containing the minimum number of independent disturbances.

A data-based approach to estimate the noise statistics is motivated by observing that the parameters characterizing the noise statistics can be viewed as 'measurable' quantities. The effect of all noises entering the plant is reflected in the measured outputs from the plant. The approach taken to estimate the noise statistics from past measured outputs and inputs is to use the expected values of the autocovariances of data at different time lags. Deviations of the model from the actual plant is accommodated in MPC by augmenting the states in the state-space model with an integrating disturbance model. The noise statistics for the integrated disturbance part of the state estimator is also estimated from data using autocovariances. In this dissertation work, a technique based on autocovariances in data is developed and applied to industrial examples. An extension of the developed technique to nonlinear models is also developed.

For illustration, we present the simplest optimal state estimator for the linear model. The optimal state estimator is the classical Kalman filter for unconstrained models. The optimal state estimates are given by the following equation where the symbol $\hat{\cdot}$ represents an estimate:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + AL_o(y_k - C\hat{x}_k)$$

in which, L_o is the optimal filter gain given by solving the following equations:

$$P_o = AP_oA^T - AP_oC^T(CP_oC^T + R_v)^{-1}CP_oA^T + GQ_wG^T$$

$$L_o = P_oC^T(CP_oC^T + R_v)^{-1}$$

The estimate error covariance is $P_o = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$. Thus to calculate L_o information about the covariances Q_w and R_v is needed. In the absence of this information the covariances are set heuristically and any suboptimal filter gain L is set using *ad-hoc* techniques to get reasonable performance from the closed-loop controller. The techniques developed in this work will be used to estimate $\hat{Q}_w$ and $\hat{R}_v$ from routine operating data to use in the above Kalman filter equations.

1.1 Dissertation Overview

The subsequent chapters in this dissertation are organized as follows:

Chapter 2:

A review of the current and past literature on data-based techniques to estimate noise statistics is presented in this chapter. The advantages of the formulation presented in this dissertation work over other correlation based techniques in the literature is shown through an illustrative example.

Chapter 3:

The formulation of the Autocovariance Least-Squares (ALS) technique is presented in

this chapter. Necessary and sufficient conditions are proved for the uniqueness of the ALS optimization. Semidefinite constraints are added to the optimization to ensure physical constraints on the covariances and a modification of the ALS technique is presented to estimate the number of independent disturbances from data.

Chapter 4:

This chapter is of value to an industrial practitioner interested in implementing the ALS technique developed in Chapter 3. In this chapter issues involving practical implementation of the ALS technique is presented along with mathematical simplications to improve the computation times. A derivation of the weighting of the ALS objective to ensure minimum variance in the estimates is also presented along with an example.

Chapter 5:

The implementation of the ALS technique with Model Predictive Control (MPC) is described. Integrating disturbance models in MPC ensure offset-free control in the presence of unmodelled disturbances. It is shown that the ALS technique estimates the covariances to give optimal closed-loop controller performance even when the source of the integrating disturbances are misassigned in the model.

Chapter 6:

The connections between maximum likelihood estimation (MLE) and the ALS technique

is established here. The MLE estimation for the covariances is shown to be an equivalent ALS optimization with the least-squares objective weighted appropriately.

Chapter 7:

The main industrial applications of the ALS technique are presented. The extension of the ALS technique to nonlinear models and an application to industrial blending drum operating data given by *ExxonMobil Chemical Company* is shown. The ALS technique is also applied to specify the estimator gains for the disturbance model structure used in *Shell*'s multivariable optimization control (SMOC). Further, the connections between a nonlinear state estimator, called the Implicit Dynamic Feedback (IDF), used at *ExxonMobil Chemical Company* and the integrating input disturbance models in MPC is described.

Chapter 8:

Nonlinear state estimation is a challenge due to the increased complexity of the models. A hybrid technique combining optimization based state estimation (moving horizon estimation) and sampling approaches to state estimation (particle filters) is described in this chapter. The hybrid estimation has the advantage of being robust to data outliers, bad initial guesses for the states and unmodelled disturbances in addition to being fast and implementable in real time.

Chapter 9:

This chapter summarizes the main contributions in this dissertation work and outlines recommended future work.

Chapter 2

Literature Review [1]

Most of the literature on estimating the noise statistics from data fall under the category of adaptive Kalman filtering. The most cited techniques in this category are based on correlations in data. The Autocovariance Least-Squares (ALS) technique presented in this work is also a correlation based technique. The literature on adaptive Kalman filtering is restricted to linear systems and open-loop data. In addition, industrial applications of adaptive Kalman filtering are lacking in the literature, which we believe is due to the omission of physical constraints such as semidefiniteness of the covariances and thus reducing the practical impact of these techniques. We attempt in this dissertation work to present a technique that is implementable in practical applications and also present rigorous mathematical proofs on the applicability of the techniques.

In the rest of this chapter, we review other literature on estimating covariances from data. In Section 2.2, we show some preliminary results to motivate the ALS technique compared to others in the literature. Since the literature is concentrated on linear

[1] **Portions of this chapter appear in Odelson et al. (2006b)**

systems, we refer to the following linear state-space model in the rest of this chapter:

$$x_{k+1} = Ax_k + Bu_k + Gw_k$$

$$y_k = Cx_k + v_k$$

with notations defined in Chapter 1.

2.1 Review of Previous Work

A literature review of the previous works to estimate noise covariances as parameters from data, identifies three major techniques: subspace ID, maximum likelihood estimation (MLE) techniques and correlation based techniques.

2.1.1 Subspace Identification Techniques

Subspace ID is one of the most popular techniques used to estimate the system parameters. The popularity of the subspace ID methods stems from the fact that the calculations are based on simple projections of matrices containing readily available output and input data. Subspace ID methods (SIMs) are however suboptimal as compared to the traditional prediction error methods (PEMs) for model parameter estimation. The downside of PEMs is that they require special parametrization and are nonlinear for multivariable systems. Hence, there are no guarantees for the global optimum or for the convergence of the optimization algorithms.

Starting with the state-space formulation in the innovations form:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + AL\mathcal{Y}_k \\ y_k &= Cx_k + \mathcal{Y}_k \end{aligned} \tag{2.1}$$

where, L is the estimator gain and $\mathcal{Y}_k = (y_k - C\hat{x}_k)$, subspace ID techniques estimate the system matrices A, B, C, L and the covariance of the innovations given by $\text{cov}(\mathcal{Y}_k) = CPC^T + R_v$ with P defined as the estimate error covariance.

Van Overschee and De Moor (1995) show that the subspace ID algorithms are based on an unifying theme. Some useful subspace ID algorithms were given by Van Overschee and De Moor (1994); Larimore (1996, 1994); Viberg (1995) and Verhaegen (1994). The above algorithms however give biased estimates for systems having inputs dependent on the past outputs. Of recent interest Li and Qin (2001) have used principal component analysis (PCA) for subspace ID using closed-loop data.

None of these algorithms however, estimate the noise covariances Q_w and R_v. The Kalman filter gain L is estimated directly and the estimates $\text{cov}(\hat{x}_{k+1} - A\hat{x}_k)$ and $\text{cov}(y_k - C\hat{x}_k)$ are used instead of $\hat{Q}_w$ and $\hat{R}_v$. The state estimator is also restricted to the unconstrained Kalman filter. The covariances Q_w, R_v are needed for advanced state estimation techniques like the Moving Horizon Estimator (MHE) that can then be used with a more general class of constrained or nonlinear systems. Estimating the covariances in comparison to gain L, also has the advantage of monitoring the noise entering the plant.

Another algorithm developed independent of the subspace ID techniques but

having an approach similar to subspace ID is the Observer/Kalman filter ID (OKID) developed by Juang et al. (1993) and Juang and Phan (1994). The difference from subspace ID techniques is that the OKID is applicable to closed-loop data. OKID works with closed-loop data because the algorithm does not use projections of data matrices into the future.

The stochastic part of the model (after subtracting the inputs) can be written as (Juang et al., 1993):

$$\begin{aligned} \hat{x}_{k+1} &= (A - ALC)\hat{x}_k + AL y_k \\ y_k &= C\hat{x}_k + y_k \end{aligned} \tag{2.2}$$

Since the optimal filter gain L provides a stable estimator, we have $(A - ALC)^p \approx 0$ for some $p > 0$. The finite output data of length N_d can then be expressed as:

$$\underbrace{\begin{bmatrix} y_1^T \\ y_2^T \\ \dots \\ y_{N_d}^T \end{bmatrix}^T}_{\mathbf{y}} = \underbrace{\begin{bmatrix} CAL & \dots & C(A-ALC)^{p-1}AL \end{bmatrix}}_{\alpha} \underbrace{\begin{bmatrix} y_0 & y_1 & y_2 & \dots & y_{N_d-1} \\ 0 & y_0 & y_1 & \dots & y_{N_d-2} \\ 0 & 0 & y_0 & \vdots & \vdots \\ 0 & 0 & 0 & \dots & y_{N_d-p} \end{bmatrix}}_{\mathbf{Y}} + \underbrace{\begin{bmatrix} y_1^T \\ y_2^T \\ y_3^T \\ \dots \\ y_{N_d}^T \end{bmatrix}^T}_{\epsilon} \tag{2.3}$$

Or in more succinct form as,

$$\mathbf{y} = \alpha \mathbf{Y} + \epsilon \tag{2.4}$$

The OKID algorithm then estimates the optimal L using the following steps. For details the reader is referred to Juang (1994) and Juang et al. (1993).

- Step 1: Estimate $\hat{\alpha}$ by solving a least-squares problem starting with Equation 2.4: $\hat{\alpha} = \mathbf{y}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}$
- Step 2: Re-parametrize $\hat{\alpha}$ and form a new least-squares problem in AL
- Step 3: Solve the second least-squares problem to get an estimate of $\hat{L}$

As opposed to the ALS technique, which solves a single least-squares optimization, the extra steps make the OKID algorithm suboptimal.

An optimization approach to finding the Kalman filter gain L is to minimize the mean square error of the estimate error (Åström, 1970; Gelb, 1974).

$$\min_L E[(x_k - \hat{x}_k)^T S(x_k - \hat{x}_k)] \tag{2.5}$$

in which S is any semidefinite matrix. If $S = I_n$, then Equation 2.5 simplifies to the Riccati equation. If S is chosen to be CC^T, then 2.5 becomes a parametric optimization, minimizing the mean squared output error i.e. minimizing $E[\mathcal{Y}_k^T\mathcal{Y}_k]$ or $E[\text{Tr}\,(\mathcal{Y}_k\mathcal{Y}_k^T)]$. Using properties of the Frobenius norm, an equivalent objective is $\|\epsilon\|_F^2$ in the limit of large N_d (the notation $\|\cdot\|_F$ is used to denote the Frobenius norm).

Given Equation 2.4, the optimization approach in Equation 2.5 then becomes:

$$\min_L \|\mathbf{y} - \alpha\mathbf{Y}\|_F^2 \tag{2.6}$$

Equation 2.6 however is a nonlinear optimization in L, as seen in Equation 2.3 and hence, the solution of optimizing algorithms depends on the initial guess. The ALS optimization on the other hand is convex and the convergence of the algorithms is guaranteed for all initial guesses.

2.1.2 Maximum Likelihood Estimation and Grey-box Modelling

For the state-space model:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Gw_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{2.7}$$

the known parameters are the matrices A, B, G, C and the unknown parameters are the covariances Q_w and R_v for the state noise w_k and measurement noise v_k respectively. With a finite set of the outputs available, maximum likelihood estimation (MLE) techniques aim to maximize the likelihood function. The innovations form of the likelihood function is given as (Shumway and Stoffer, 2000):

$$-2\ln L_Y(\theta) = \sum_{k=1}^{N_d} \log|\Sigma_k(\theta)| + \sum_{k=1}^{N_d} \mathcal{Y}_k(\theta)^T \Sigma_k(\theta)^{-1} \mathcal{Y}_k(\theta) \tag{2.8}$$

in which $L_Y(\theta)$ is the likelihood and $\Sigma_k = CP_kC^T + R_v$. θ is the vector containing the parameters (A, B, C, G, Q_w, R_v).

If the only unknowns of interest are the covariances and the rest of the parameters are known with reasonable confidence, then the estimation procedure follows an iterative scheme to minimize the objective in Equation 2.8 over the covariances Q_w, R_v with no guarantees of convergence. Note that the innovations $\mathcal{Y}_k$ are dependent on the parameter θ and the likelihood function is a highly nonlinear and complicated function of θ. Newton-Raphson based algorithms (Shumway and Stoffer, 1982) and other algorithms based on the expectation-maximization (EM) (Dempster et al., 1977) have been developed in the literature to solve the above problem. All these algorithms however, are iterative in nature and convergence is not proved.

A second class of problems falling under maximum likelihood estimation procedures is using grey-box models. Grey-box models combine physical information with a nonlinear state-space model yielding fewer parameters having physical meaning. Bohlin and Graebe (1995) provide a Matlab toobox to estimate parameters in grey-box models. The algorithm and the toolbox was further developed and improved by Kristensen et al. (2004a,b). Their algorithm is also based on the maximization of the likelihood function. Other than these references there has been little literature on the identification of grey-box models. The algorithm in the Matlab toolbox involves a nonlinear optimization of the parameters A, C, G and the noise covariances Q_w, R_v in the model and a recursive calculation of the innovations through an Extended Kalman Filter (EKF) for each guess of the parameters. It can be easily seen that as the number of states increases the size of the vector of unknowns θ becomes large and the estimation procedure becomes complicated and unreliable.

For the purpose of evaluation of the MLE techniques in the literature, the deterministic parts of the model i.e. A, G and C were fixed and the likelihood function was optimized with $\theta = [(Q_w)_s, (R_v)_s]$. The ML estimation algorithm was tested by using a software package called Continuous Time Stochastic Model (CTSM) developed by Kristensen et al. (2004a,b) and available for download at `http://www.imm.dtu.dk/ctsm/download.html`. The estimation of the covariances was highly dependent on the initial guesses and failed to give positive semidefinite covariances for good initial guesses. The estimated covariances were positive semidefinite only when the covariance ma-

trices were constrained to be diagonal. For the same set of data the Autocovariance Least-Squares (ALS) procedure gave excellent results and the technique was executed within short computational times.

2.1.3 Correlation Techniques

Correlation techniques are based on the idea that once the deterministic part of a plant is modelled accurately, the residuals from the deterministic part then carry information about the noises entering the plant. These residuals (or innovations) can then be correlated with each other to extract information about the covariance of the disturbances entering the plant. The main proponent of this idea was Mehra (1970, 1971, 1972) and adapted by many others (Neethling and Young, 1974; Isaksson, 1987; Carew and Bélanger, 1973; Bélanger, 1974; Noriega and Pasupathy, 1997). The ALS technique is also a correlation based technique and initial work was presented in Odelson et al. (2006b) and Odelson (2003). The ALS technique offers significant advantages over other techniques in the literature. The ALS procedure solves a single least-squares problem while the other techniques estimate the covariances in two steps. Solving a single least-squares problem leads to smaller variance in the estimates as opposed to using two steps. The correlation methods in the literature also, do not impose semidefinite constraints on the covariances leading to estimates that is not positive semidefinite for finite sets of data.

Other literature not reviewed above are summarized below. An adaptive scheme

to calculate the Kalman filter gain using past innovations has been suggested by Hoang et al. (1994, 1997). However, this method takes many samples to converge and fails to achieve the linear minimum-variance estimate (LMVE) when deterministic disturbances enter the plant input or output. A recent method which minimizes the output residuals in the least-squares sense (Juang et al., 1993; Juang and Phan, 1994; Juang, 1994) was found to be inadequate as the method does not find the covariances, but finds the filter gain directly. An additional step of fitting a moving average model to the suboptimal filter residuals (Chen and Huang, 1994) is not amenable for multivariable systems and also fails given integrated white noise disturbances. Methods for estimating the state when there are disturbances other than white noise in the plant usually involves two steps. The disturbance model is identified first using the internal model principle and then the disturbance is rejected using some arbitrary pole placement for the observer that may be far from being optimal (Palaniswami and Feng, 1991; Xu and Shou, 1991). Maryak et al. (2004) propose bounds for the Kalman filter gain for models with unknown noise distributions. Sparks and Bernstein (1997) also use the internal model principle to represent deterministic and sinusoidal disturbances, and their approach involves supercompensators and dynamic compensators (Davison and Ferguson, 1981; Davison and Goldenberg, 1975).

2.2 Comparison to other Correlation based Methods

In this section we present some preliminary results showing the advantages of using the ALS technique against those in the literature. Chapter 3 provides more details on the ALS technique formulation and the technical details.

As mentioned in Section 2.1.3, Mehra (1972); Neethling and Young (1974); Carew and Bélanger (1973) and Bélanger (1974) employ a three-step procedure to estimate (Q_w, R_v) from data. Starting with the innovations form for the state-space equation, the steps are: (i) Solve a least-squares problem to estimate PC^T from the estimated autocovariances (P is the estimate error covariance). (ii) Use the estimated PC^T to solve for R_v using the zero lag autocovariance estimate from data. (iii) Solve a least-squares problem to estimate Q_w from the estimated PC^T and R_v.

We offer two criticisms of the classic Mehra approach. The first comment concerns the conditions for uniqueness of $(\hat{Q}_w, \hat{R}_v)$ in Mehra's approach. These conditions were stated (without proof) as

1. (A, C) observable
2. A full rank
3. The number of unknown elements in the Q_w matrix, $g(g+1)/2$, is less than or equal to np (n is the number of states and p is the number of outputs)

These conditions were also cited by Bélanger (1974). We can easily find counterexamples for these conditions as reported in Odelson et al. (2006b).

The second comment concerns the large variance associated with Mehra's method. This point was first made by Neethling and Young (1974), and seems to have been largely overlooked. First, step (ii) above is inappropriate because the zero-order lag autocovariance estimate, $(CPC^T + R_v)$ is not known perfectly. Second, breaking a single-stage estimation of Q_w and R_v into two stages by first finding PC^T and R_v and then using these estimates to estimate Q_w in steps (i) and (iii) also increases the variance in the estimated Q_w. To quantify the size of the variance inflation associated with Mehra's method, consider the following example, which has a well-conditioned observability matrix:

$$A = \begin{bmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad C = \begin{bmatrix} 0.1 & 0.2 & 0 \end{bmatrix}$$

Data are generated using noise sequences with covariance $Q_w = 0.5, R_v = 0.1$. The innovations are calculated with a filter gain corresponding to incorrect noise variances $Q_w = 0.2$ and $R_v = 0.4$. Mehra's method and the single-step ALS method are run using $N_d = 1000$ data points, $N = 15$. The simulation is repeated 200 times to illustrate the mean and variances of the estimators.

In Figure 2.1, the estimates of (Q_w, R_v) using Mehra's method are plotted. The variance of the estimates is large, and many of the estimates are negative, which is unphysical. In Figure 2.2, the ALS estimates of (Q_w, R_v) are plotted, on much tighter axes. The variance of the ALS estimates is much smaller than in Mehra's method; and none of the estimates are negative. Note that Neethling and Young (1974) discuss other

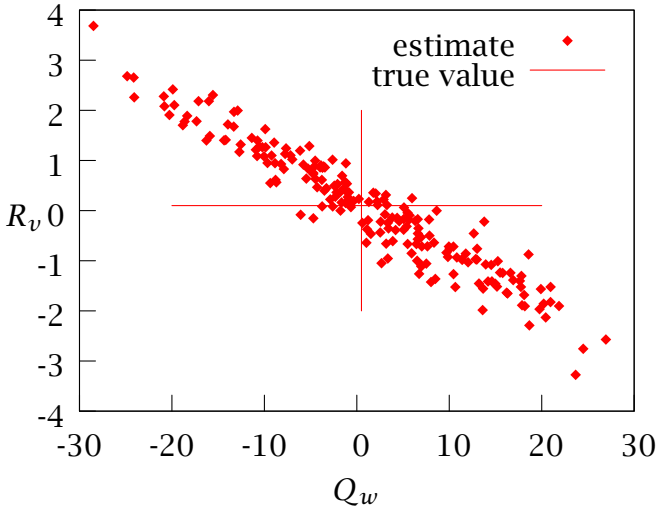


Figure 2.1: Estimates of Q_w and R_v using Mehra's method

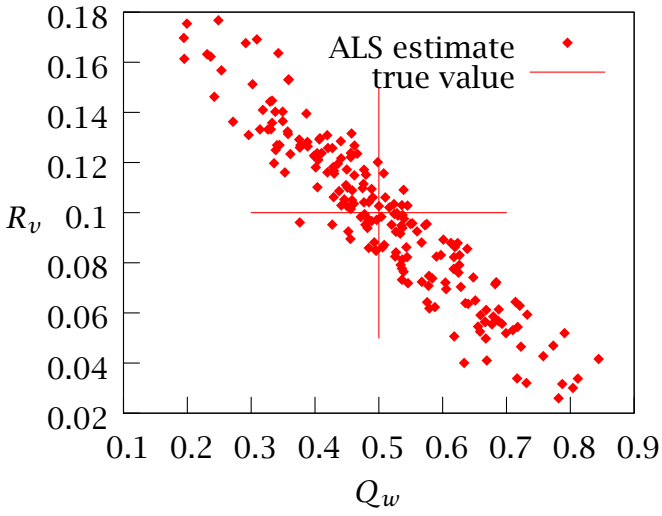


Figure 2.2: Estimates of Q_w and R_v using proposed ALS method. Notice the axes have been greatly expanded compared to Fig. 2.1

examples with behavior similar to this one.

When dealing with a small sample of measurements or significant plant/model error, the estimate of the covariances using ALS may not be positive semidefinite, even

though the variance of the estimate may be smaller than the three-step procedure. Such estimates are physically meaningless. Most of the literature for estimating covariances does not address this issue. A recent *ad-hoc* method of imposing positive semidefiniteness on the estimates of R_v only given in Noriega and Pasupathy (1997). The ALS technique presented in this work as detailed in Chapter 3 impose the semidefinite constraints directly. The semidefinite constraints are convex in Q_w, R_v and the optimization is in the form of a semidefinite programming (SDP) problem (Vandenberghe and Boyd, 1996). A simple path following algorithm based on Newton steps provides a simple and efficient method to find the global optimum. Details of other semidefinite optimization algorithms in the literature can be found in Wolkowicz et al. (2000, chap. 10).

Chapter 3

The Autocovariance Least-Squares (ALS) Technique with Semidefinite Programming (SDP) [1]

We start with the discrete linear time-invariant state-space model given in Chapter 1:

$$x_{k+1} = Ax_k + Bu_k + Gw_k \tag{3.1a}$$

$$y_k = Cx_k + v_k \tag{3.1b}$$

in which x_k, u_k and y_k are the state, input and output of the system at time t_k. The dimensions of the system matrices are $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $G \in \mathbb{R}^{n\times g}$ and $C \in \mathbb{R}^{p\times n}$. The noises corrupting the state and the output ($w_k \in \mathbb{R}^g$ and $v_k \in \mathbb{R}^p$) are modelled as zero-mean Gaussian noise sequences with covariances Q_w and R_v respectively. The noises w_k and v_k are assumed to be statistically independent for simplicity. The case

[1] **Portions of this chapter are to appear in Rajamani and Rawlings (September, 2007)**

where w_k and v_k are dependent can be handled as shown in Åkesson et al. (2007). The optimal filtering or state estimation for the model given in Equations 3.1a, 3.1b when there are no constraints on the input and the state is given by the classical Kalman filter (Kalman and Bucy, 1961). If the Gaussian assumption is relaxed, the Kalman filter is the still the optimal filter among the class of all linear filters (Goodwin and Sin, 1984; Anderson and Moore, 1979).

The G matrix shapes the disturbance w_k entering the state. Physical systems often have only a few independent disturbances which affect the states. This implies a tall G matrix with more rows than columns. All of the correlation based techniques in the literature described in Chapter 2, Section 2.1.3, assume that the disturbance structure as given by the G matrix is known. In the absence of any knowledge about G an assumption that $G = I$ is often made, which implies that an independent disturbance enters each of the states. This type of independence of the disturbances is unlikely for physical reasons. There exists no technique in the literature to estimate the structure of the disturbances entering the state, which we do in this chapter. Throughout we assume complete knowledge about A, B, C and treat the stochastic part of the model as the only unknowns.

The rest of the chapter is organized as follows: In Section 3.1 we give some mathematical preliminaries that are required to understand the rest of the chapter. Section 3.2 gives the formulation of the Autocovariance Least-Squares (ALS) technique. The main contributions of this chapter are then presented in Sections 3.3 and 3.4. Sim-

ple mathematical conditions to check for uniqueness of the covariance estimates are proved in Section 3.3 and the results used in the remaining sections. In Section 3.4 we estimate the noise shaping matrix G from data using Semidefinite Programming (SDP). The G matrix contains information about the disturbance structure and the number of independent disturbances affecting the state is equal to the number of columns in G.

3.1 Background

Assumption 3.1. *We assume that the pair* (A, C) *is observable*

We use the notation $\hat{x}_k$ to denote any estimate of the state x_k. If $L \in \mathbb{R}^{n \times p}$ is any arbitrary, stable filter gain, then the state estimates are calculated recursively as:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + AL(y_k - C\hat{x}_k) \tag{3.2}$$

When the system is unconstrained, the optimal state estimator is the Kalman filter. For the Kalman filter the filter gain L_o is calculated by solving the Riccati equation (also presented in Chapter 1):

$$\begin{aligned} P_o &= AP_oA^T - AP_oC^T(CP_oC^T + R_v)^{-1}CP_oA^T + GQ_wG^T \\ L_o &= P_oC^T(CP_oC^T + R_v)^{-1} \end{aligned} \tag{3.3}$$

The optimal estimate error covariance is $P_o = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$ calculated as above and requires knowledge about the covariances Q_w and R_v.

Given some arbitrary (stable, perhaps suboptimal) initial estimator L, we can write the evolution of the state estimate error $\varepsilon_k = x_k - \hat{x}_k$ by subtracting Equation 3.2 from

3.1a and substituting 3.1b:

$$\varepsilon_{k+1} = \underbrace{(A - ALC)}_{\bar{A}} \varepsilon_k + \underbrace{\begin{bmatrix} G & -AL \end{bmatrix}}_{\bar{G}} \begin{bmatrix} w_k \\ v_k \end{bmatrix} \tag{3.4}$$

$$y_k = C\varepsilon_k + v_k$$

in which y_k are the L-innovations defined as $y_k \triangleq y_k - C\hat{x}_k$. Note that the L-innovations are uncorrelated in time if the initial state estimator L is optimal (i.e. $L = L_o$) (Anderson and Moore, 1979). We use the term L-innovations to distinguish them from the optimal innovations obtained by using the optimal state estimates.

Assumption 3.2. *The L-innovations data $\{y_1, \cdots y_{N_d}\}$ used in the techniques described in this chapter are obtained after the system has reached steady state and any initial transience is neglected since $\bar{A}$ is stable*

Given a set of steady state L-innovations data $\{y_1, \cdots y_{N_d}\}$, we want to form a weighted least-squares problem in the unknown disturbance covariances, GQ_wG^T, R_v. One of the motivations behind using a least-squares approach is to avoid a complicated nonlinear approach required for techniques involving maximum likelihood estimation eg. Shumway and Stoffer (1982).

In the subspace ID literature (Gevers, 2006; Van Overschee and De Moor, 1994, 1995; Viberg, 1995; Juang and Phan, 1994; Qin et al., 2005), the identification procedures estimate the model and the stochastic parameters starting with the model in the

innovations form, which is Equation 3.2 rewritten as:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + AL_o e_k \tag{3.5a}$$

$$y_k = C\hat{x}_k + e_k \tag{3.5b}$$

Here e_k are the optimal innovations (as opposed to the L-innovations) and hence uncorrelated in time. The estimation of the system matrices $\hat{A}, \hat{B}, \hat{C}$ is carried out along with the optimal Kalman filter gain $\hat{L}_o$, where the $\hat{\cdot}$ symbol denotes an estimate.

Notice the difference between Equations 3.5a,3.5b and Equations 3.1a, 3.1b. If the subspace ID techniques are used to identify only the stochastic parameters then the disturbance covariances as identified as $A\hat{L}_o S\hat{L}_o^T A^T$ instead of $GQ_w G^T$ for the state noise and S instead of R_v for the measurements, where S is the covariance of e_k given by:

$$S = CP_o C^T + R_v$$

where, P_o is defined in Equation 3.3.

Remark 3.1. *As shown above, subspace ID techniques estimate a different set of covariances than G, Q_w, R_v. The aims of subspace ID are different and the estimates of the stochastic parameters are simply used to compute the optimal estimator gain. Finding the covariance parameters affecting the system (G, Q_w, R_v) on the other hand provides more flexibility in the choice of the state estimator. For example we may wish to employ a constrained, nonlinear moving horizon estimator (Rao et al., 2001). In addition*

estimating G, Q_w, R_v gives a more informative handle to monitor the disturbances than monitoring changes in the optimal estimator gain.

Also see Remark 3.2 for requirements about exciting inputs in subspace ID techniques.

3.2 The Autocovariance Least-Squares (ALS) Technique

Following the derivation along the lines of Odelson et al. (2006b), we use Equation 3.4 to write the following expectation of covariances:

$$E(y_k y_k^T) = CPC^T + R_v \tag{3.6}$$

$$E(y_{k+j} y_k^T) = C\bar{A}^j PC^T - C\bar{A}^{j-1} ALR_v \quad j \geq 1 \tag{3.7}$$

which are independent of k because of our steady state assumption. Again using Equation 3.4 we note that P satisfies the Lyapunov equation:

$$P = \bar{A}P\bar{A}^T + \underbrace{\begin{bmatrix} G & -AL \end{bmatrix}}_{\bar{G}} \underbrace{\begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}}_{\bar{Q}_w} \bar{G}^T \tag{3.8}$$

In Odelson et al. (2006b) the autocovariance matrix was defined as:

$$\mathcal{R}(N) = E \begin{bmatrix} y_k y_k^T & \cdots & y_k y_{k+N-1}^T \\ \vdots & \ddots & \vdots \\ y_{k+N-1} y_k^T & \cdots & y_k y_k^T \end{bmatrix} \tag{3.9}$$

where N is the number of lags. To avoid redundant definition of the lagged covariances, here we use only the first block column of the autocovariance matrix $\mathcal{R}_1(N)$:

$$\mathcal{R}_1(N) = E\begin{bmatrix} y_k y_k^T \\ \vdots \\ y_{k+N-1} y_k^T \end{bmatrix} \tag{3.10}$$

The formulation of the ALS technique for the full autocovariance matrix is given in Appendix 3.6.4.

Using Equations 3.6, 3.7 and 3.8, we can write the $\mathcal{R}_1(N)$ as:

$$\mathcal{R}_1(N) = \mathcal{O}PC^T + \Gamma R_v \tag{3.11}$$

in which

$$\mathcal{O} = \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-1} \end{bmatrix} \quad \Gamma = \begin{bmatrix} I_p \\ -CAL \\ \vdots \\ -C\bar{A}^{N-2}AL \end{bmatrix} \tag{3.12}$$

The single column block development of the ALS technique as above is preferred over the use of the full $\mathcal{R}(N)$ matrix as in Odelson et al. (2006b) due to the simpler formulation when using only $\mathcal{R}_1(N)$.

In this result and those to follow, we employ the standard definitions and properties of the Kronecker product, Kronecker sum and the direct sum (Steeb (1991); Graham (1981, chap. 2) and Van Loan (2000)). If use the 's' subscript to denote the column-wise stacking of the matrix into a vector, a useful Kronecker product result is $(AXB)_s = (B^T \otimes A)X_s$ (here $\otimes$ is the standard symbol for the Kronecker product).

We then stack Equation 3.11 and use the stacked form of Equation 3.8 to substitute out P:

$$\begin{aligned} b = (\mathcal{R}_1(N))_s = &[(C \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}](GQ_wG^T)_s \\ &+ [(C \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(AL \otimes AL) \\ &+ (I_p \otimes \Gamma)](R_v)_s \end{aligned} \tag{3.13}$$

Now that we have Equation 3.13, we use the ergodic property of the L-innovations to estimate the autocovariance matrix $\mathcal{R}_1(N)$ from the given set of data (Jenkins and Watts, 1968):

$$\widehat{E[y_k y_{k+j}^T]} = \frac{1}{N_d - j} \sum_{i=1}^{N_d - j} y_i y_{i+j}^T \tag{3.14}$$

If $\{y_1, y_2, \cdots y_{N_d}\}$ are the set of L-innovations calculated from data as given by Equation 3.4, and N is the window size used for the autocovariances then we define the matrix $\mathbb{Y}$ as follows:

$$\mathbb{Y} \triangleq \begin{bmatrix} y_1 & y_2 & \cdots & y_{N_d-N+1} \\ y_2 & y_3 & \cdots & y_{N_d-N+2} \\ \vdots & \vdots & \vdots & \vdots \\ y_N & y_{N+1} & \vdots & y_{N_d} \end{bmatrix} \tag{3.15}$$

$\mathbb{Y} \in \mathbb{R}^{\tilde{p} \times \tilde{n}}$ where, $\tilde{n} \triangleq N_d - N + 1$ and $\tilde{p} \triangleq Np$. Using Equation 3.14, the estimate $\widehat{\mathcal{R}_1}(N)$ is then given by:

$$\widehat{\mathcal{R}_1}(N) = \frac{1}{N_d - N + 1} \mathbb{Y}\mathbb{Y}_1^T \tag{3.16}$$

and $\hat{b} = (\widehat{\mathcal{R}_1}(N))_s$. Here, $\mathbb{Y}_1$ is the first row block of $\mathbb{Y}$ also given by:

$$\mathbb{Y}_1 = \underbrace{\begin{bmatrix} I_p & 0 & \cdots & 0 \end{bmatrix}}_{\mathbb{E}} \mathbb{Y} \tag{3.17}$$

Given the linear relation in Equation 3.13 and the estimate $\hat{b}$ from Equation 3.16, we can formulate the following positive semidefinite constrained least-squares problem in the symmetric elements of the covariances GQ_wG^T, R_v:

$$\begin{aligned} \Phi = \min_{GQ_wG, R_v} \quad & \left\| \mathcal{A} \begin{bmatrix} \mathcal{D}_n (GQ_wG^T)_{ss} \\ (R_v)_{ss} \end{bmatrix} - \hat{b} \right\|_W^2 \\ \text{subject to,} \quad & GQ_wG^T, R_v \succeq 0, \quad R_v = R_v^T \end{aligned} \tag{3.18}$$

The matrix $W > 0$ denotes the weighting of the least-squares objective. More details on the derivation of an optimal W are given in Chapter 4, Section 4.1. Here, we introduce the notation of $(R_v)_{ss}$ to denote the column-wise stacking of only the symmetric $p(p+1)/2$ elements of the matrix R_v (eliminating the supra-diagonal elements). More explicitly there exists an unique matrix $\mathcal{D}_p \in \mathbb{R}^{p^2 \times \frac{p(p+1)}{2}}$ called the *duplication matrix* (Magnus and Neudecker, 1999, p. 49) containing ones and zeros that gives the relation $(R_v)_s = \mathcal{D}_p (R_v)_{ss}$.

Using Equation 3.13, we can then write $\mathcal{A}$ explicitly as:

$$\begin{aligned} \mathcal{A} =& [\mathcal{A}_1 \quad \mathcal{A}_2] \\ \mathcal{A}_1 =& [(C \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}] \\ \mathcal{A}_2 =& [(C \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(AL \otimes AL) \\ & + (I_p \otimes \Gamma)]\mathcal{D}_p \end{aligned} \tag{3.19}$$

where, the *duplication matrix* $\mathcal{D}_p$ is included to ensure symmetry in the covariance estimates.

The estimation method in Equation 3.18 is referred to as the Autocovariance Least-Squares (ALS) technique in the sequel. A recent application of the ALS technique was presented in Zhuang et al. (2007a,b). The ALS technique can also be used to estimate the optimal filter gain when there are integrating disturbance models in model predictive control (Chapter 5 and Rajamani et al. (2006)). One major advantage of the semidefinite formulation of the ALS technique in Equation 3.18 is that it can be easily solved by available semidefinite programming solvers (Wolkowicz et al., 2000; Vandenberghe and Boyd, 1996; Boyd et al., 1994).

Remark 3.2. *A significant advantage of using the ALS technique and the modifications presented in the rest of this chapter over other identification techniques is the use of only steady state data in the calculations. This means that unlike other identification techniques there is no requirement for exciting inputs to be applied to the system.*

3.3 Conditions for Uniqueness

In this section, we assume that the G matrix is a known $\mathbb{R}^{n\times g}$ matrix. Without loss of generality we can also assume G to be of full column rank. If G is not full column rank then a new matrix $\tilde{G}$ can be defined with its columns independent and such that $\tilde{G}\tilde{G}^T = GG^T$.

We next derive simple conditions for uniqueness for the ALS problem with Q_w, R_v as unknowns and a known G. In the rest of this section we also assume that the weighting for the norm in the objective function is $W = I$.

$$\Phi(Q_w, R_v) = \min_{Q_w, R_v} \left\| \tilde{\mathcal{A}} \begin{bmatrix} (Q_w)_{ss} \\ (R_v)_{ss} \end{bmatrix} - \hat{b} \right\|^2 \tag{3.20}$$

$$\text{s.t.} \quad Q_w, R_v \geq 0, \quad R_v = R_v^T, \quad Q_w = Q_w^T$$

where, $\tilde{\mathcal{A}} = \begin{bmatrix} \mathcal{A}_1 (G \otimes G) \mathcal{D}_g & \mathcal{A}_2 \end{bmatrix}$

Lemma 3.1. *The solution of the optimization in Equation 3.20 exists for all $\hat{b}$. The solution is unique if and only if $\tilde{\mathcal{A}}$ in Equation 3.20 has full column rank.*

Proof. Since $\hat{b}$ is a finite value calculated from data, given the bound $B = \Phi(I_g, I_p)$ from Equation 3.20, the inequality $\{Q, R | \Phi(Q, R) \leq B\}$ is a bounded and closed, hence compact, set (Haaser and Sullivan, 1991, Heine-Borel Theorem, p. 85). The set $\{Q, R | Q \geq 0, R \geq 0, Q = Q^T, R = R^T\}$ is the set of symmetric positive semidefinite matrices and hence is a closed set (Wolkowicz et al., 2000, p. 69). The intersection of a compact set and a closed set is compact (Haaser and Sullivan, 1991, p. 80). The minimization of a continous function (Φ) on a compact set achieves its minimum on the set (Sundaram, 1996, Weierstrass Theorem, p. 90). Therefore, the solution to Equation 3.20 exists for any set of finite measurements.

To prove uniqueness, we see that $\tilde{\mathcal{A}}$ having full column rank guarantees the objective function in Equation 3.20 to be strictly convex. The positive semidefinite constraints in Equation 3.20 are also convex (Vandenberghe and Boyd, 1996; Boyd et al.,

1994). The set $\{Q, R | \Phi(Q, R) \leq B\}$ defines a convex set, which when intersected with the semidefiniteness requirement still defines a convex set. Uniqueness then follows from the existence of a bounded solution for all $\hat{b}$ and the fact that a strictly convex objective function is subject to convex constraints (Boyd and Vandenberghe, 2004, p. 137). □

Assumption 3.3. *We assume that the state transition matrix A in non-singular. If the original A is singular, then a similarity transformation can be used to eliminate the states with zero eigenvalues and the noise covariances redefined.*

Lemma 3.2. *If (A, C) is observable and A is non-singular, then the matrix $\tilde{\mathcal{A}}$ in Equation 3.20 has a null space if and only if the matrix M defined by, $M = (C \otimes I_n)(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(G \otimes G)\mathcal{D}_g$ also has a null space, and the null space of $\mathcal{A}_1(G \otimes G)\mathcal{D}_g$ which multiplies $(Q_w)_{ss}$ in Equation 3.20 is equal to the null space of M.*

Proof of Lemma 3.2 is given in Appendix 3.6.1.

Theorem 3.1. *If (A, C) is observable and A is non-singular, the optimization in Equation 3.20 has a unique solution if and only if $dim[Null(M)] = 0$, where:*

$$M = (C \otimes I_n)(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(G \otimes G)\mathcal{D}_g$$

Proof. The proof follows from Lemmas 3.1 and 3.2. □

Corollary 3.1. *If C is full column rank (i.e. the number of sensors equal the number of states), then the optimization in Equation 3.20 is unique.*

Proof. C having full column rank implies M in Theorem 3.1 has full rank and hence an empty null space. The optimization in Equation 3.20 then gives a unique solution according to Theorem 3.1. □

3.4 The ALS-SDP method

In this section the G matrix is also assumed to be unknown in addition to the Q_w, R_v matrices. An estimation technique is presented that estimates the structure of the G matrix modelling the minimum number of independent disturbances affecting the state.

Generally a linear model of a system has many states and only a few independent disturbances corrupting these states. Any noise w_k that enters the state x_{k+1} is first scaled by the G matrix and then by the C matrix before it is measured in the output y_{k+1} (Equations 3.1a and 3.1b). It is unlikely to have information about the G matrix in most applications. Information contained in the measurements is also usually not enough to estimate a full GQ_wG^T matrix uniquely (this can be checked using Theorem 3.1). If there are fewer sensors than the states, there can be multiple covariances that generate the state noises making up the same output data (Corollary 3.1).

When G is unknown, our aim is to find the minimum rank Q (where, $Q = GQ_wG^T$). A minimum rank Q can be decomposed as follows:

$$Q = \tilde{G}\tilde{G}^T, \qquad \tilde{Q}_w = I \tag{3.21}$$

It should be noted that the choice $\tilde{Q}_w = I$ is not a binding choice for the covariance

because any other choice of $\tilde{Q}_w$ can be easily absorbed into $\tilde{G}$ by redefining $\tilde{G}_1 = \tilde{G}Q_w^{-0.5}$ so that $Q = \tilde{G}\tilde{G}^T = \tilde{G}_1 Q_w \tilde{G}_1^T$.

Having Q with minimum rank would ensure that $\tilde{G}$ has the minimum number of columns. The number of columns in the matrix G is equal to the number of independent disturbances entering the state and equal to the rank of Q. Hence, by estimating $\tilde{G}$, we get information about the minimum number of independent disturbances entering the data in addition to the disturbance structure.

Remark 3.3. *With reference to Equation 3.21, one might think that a more natural procedure would be to solve the ALS optimization in Equation 3.18 directly with G as the optimization variable and constraining Q_w instead of solving with Q and then following with the decomposition. The reason for solving with Q as the optimization variable is to avoid the nonlinearity that would be introduced if the elements of G are used as optimization variables and the extra flexibility in allowing for minimization of the rank of Q.*

In the development of the remaining results in this section , we take the weight $W = I$. The rank can be explicitly added to the objective in Equation 3.18 through a tradeoff parameter ρ multiplying the rank:

$$\Phi_* = \min_{Q,R_v} \underbrace{\left\| \mathcal{A} \begin{bmatrix} (Q)_s \\ (R_v)_s \end{bmatrix} - \hat{b} \right\|^2}_{\Phi} + \rho \text{Rank } (Q) \tag{3.22}$$

$$Q, R_v \geq 0, \quad Q = Q^T, \quad R_v = R_v^T$$

The constraints are in the form of convex Linear Matrix Inequalities (LMI) (Boyd et al., 1994; VanAntwerp and Braatz, 2000). The norm part of the objective is also convex. The rank however can take only integer values making the problem NP hard. The solution of minimizing the rank subject to LMI constraints is an open research question and current techniques are largely based on heuristics (Vandenberghe and Boyd, 1996).

Since the rank is equal to the number of nonzero eigenvalues of a matrix, a good heuristic substitute for the rank is the sum of its eigenvalues or the trace of the matrix. The trace of a matrix is also the largest convex envelope over the rank of the matrix (Fazel, 2002).

$$\text{Rank } (Q)_{\min} \geq \frac{1}{\lambda_{\max}(Q)} \text{Tr } (Q)$$

The trace of a matrix is a convex function of Q. The optimization in Equation 3.22 can be rewritten with the trace replacing the rank:

$$\Phi_1 = \min_{Q,R_v} \underbrace{\left\| \mathcal{A} \begin{bmatrix} (Q)_s \\ (R_v)_s \end{bmatrix} - b \right\|^2}_{\Phi} + \rho \text{Tr } (Q) \qquad (3.23)$$

$$Q, R_v \geq 0, \quad Q = Q^T, \quad R_v = R_v^T$$

Lemma 3.3. *Given an optimization problem in the matrix variable $X \in \mathbb{R}^{n \times n}$ with the following form:*

$$\min_X (AX_s - b)^T (AX_s - b) + \rho \, Tr \, (X)$$

$$\textit{subject to} \quad X \geq 0, \quad X = X^T$$

with the matrices A and b appropriately dimensioned, the optimization can be rewritten in the following standard primal Semidefinite Programming problem:

$$\begin{aligned} \min_{x} \quad & c^T x \\ \textit{subject to} \quad & F(x) \geq 0 \\ \textit{where} \quad & F(x) \triangleq F_0 + \sum_{i=1}^{m} x_i F_i \end{aligned}$$

with the symmetric matrices $F_0, \cdots, F_m \in \mathbb{R}^{n \times n}$ *and the vector* $c \in \mathbb{R}^m$ *chosen appropriately.*

Proof of Lemma 3.3 is given in Appendix 3.6.2.

Given the above Lemma 3.3, if we define $X = \text{diag}(Q, R_v)$ then Equation 3.23 is in the form of a Semidefinite Programming (SDP) problem with A defined accordingly. We refer to this problem as the ALS-SDP (Autocovariance Least-Squares with Semidefinite Programming) in the sequel.

Lemma 3.4. *If* $p < n$ *(i.e. number of measurements is fewer than the number of states), then the following holds for Equation 3.23:*

$$dim[Null(\mathcal{A})] \geq (n-p)(n-p+1)/2$$

Proof. The dimension condition follows by substituting $G = I$ in Lemma 3.2, noting that $(I_{n^2} - \bar{A} \otimes \bar{A})$ is full rank and using the rank condition in Hua (1990). □

Theorem 3.2. *A solution* $(\hat{Q}, \hat{R}_v)$ *to the ALS-SDP in Equation 3.23 is unique if* $dim[Null(M)] =$

0 *where,*

$$M = (C \otimes I_n)(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(G \otimes G)\mathcal{D}_g$$

and G is any full column rank decomposition of $\hat{Q} = GG^T$ *(G is an unique decomposition within an orthogonal matrix multiplication).*

Proof. The function:

$$\Phi = \left\| \mathcal{A} \begin{bmatrix} (G \otimes G)(Q_w)_s \\ (R_v)_s \end{bmatrix} - b \right\|^2$$

is the first part of the objective in Equation 3.23 and also the same as the objective in Equation 3.20. Following Theorem 3.1 and Lemma 3.1, $\dim[\text{Null}(M)] = 0$ then implies that Φ is strictly convex at the solution $Q_w = I_g, R_v = \hat{R}_v$.

The other part of the objective in Equation 3.23 i.e. Tr (Q) is linear in the variable Q and hence is also convex. The overall objective in Equation 3.23 is then strictly convex at the solution $\hat{Q}, \hat{R}_v$ when $\dim[\text{Null}(M)] = 0$. Uniqueness of the solution thus follows from minimization of a strictly convex objective subject to convex constraints (Boyd and Vandenberghe, 2004). □

The ALS-SDP method gives a feasible solution for each value of the tradeoff parameter ρ by using simple Newton-like algorithms. The choice of ρ is made from a tradeoff plot of Tr (Q) versus Φ from Equation 3.23. A good value of ρ is such that Tr (Q) is small and any further decrease in value of Tr (Q) causes significant increase in the value of Φ. This ensures that the rank of Q is minimized without significant compromise on the original objective Φ (Rajamani and Rawlings, 2006).

The matrix inequalities $Q_w \geq 0, R_v \geq 0$ can be handled by an optimization algorithm adding a logarithmic barrier function to the objective. The optimization algorithm then minimizes:

$$\Phi_1 = \min_{Q_w, R_v} \Phi + \rho \text{Tr}\,(Q) - \mu \log \begin{vmatrix} Q & 0 \\ 0 & R_v \end{vmatrix} \tag{3.24}$$

in which, μ is the *barrier parameter* and $|\cdot|$ denotes the determinant of the matrix (Nocedal and Wright, 1999, chap. 17). The log-determinant barrier is an attractive choice because it has analytical first and second derivatives. Appendix 3.6.3 lists some useful matrix derivatives arising in the optimization in Equation 3.24. As with other barrier techniques, with $\mu \to 0$, the solution to the SDP tends to the optimum solution. The following approach was used to solve the barrier augmented SDP.

1. Choose a value for the tradeoff parameter ρ
2. Iteration k=0
3. Choose a starting value of μ (say $\mu = 100$)
4. Solve the SDP and let the solution be Q_k, R_k
5. Decrease value of μ (say choose the new value as $\mu/2$)
6. Increase value of k by 1 and repeat step 4 till $\mu < 10^{-7}$
7. Check conditions in Theorem 3.2 for uniqueness. If the solution is not unique then repeat with higher value for ρ.

Other path following type of algorithms can be found in Wolkowicz et al. (2000, chap. 10). The convexity of Equation 3.24 ensures a unique termination of the minimization algorithm. The algorithm scales efficiently for large dimensional problems.

3.4.1 Example

Let the plant be simulated using the following state-space matrices.

$$A = \begin{bmatrix} 0.733 & -0.086 \\ 0.172 & 0.991 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

with noises drawn from the distributions:

$$w_k \sim N(0, 0.5), \qquad v_k \sim N(0, 1)$$

Although the data is generated by a single column G matrix, we assume G is unknown and estimate it using the ALS-SDP procedure.

The results from the new ALS-SDP are shown in Figures 3.1 and 3.2. The plots show that choice of $\rho = 0.31$ is where the Tr (Q) is the minimum with no significant change in Φ. Also, the rank(Q) at $\rho = 0.31$ is 1, which is the number of independent disturbances entering the state in the simulated data (columns of G).

Also the estimated disturbance structure and covariances using $\rho = 0.31$ is:

$$\hat{Q} = \begin{bmatrix} 0.449 & 0.249 \\ 0.249 & 0.138 \end{bmatrix}, \quad \hat{R}_v = 0.99$$

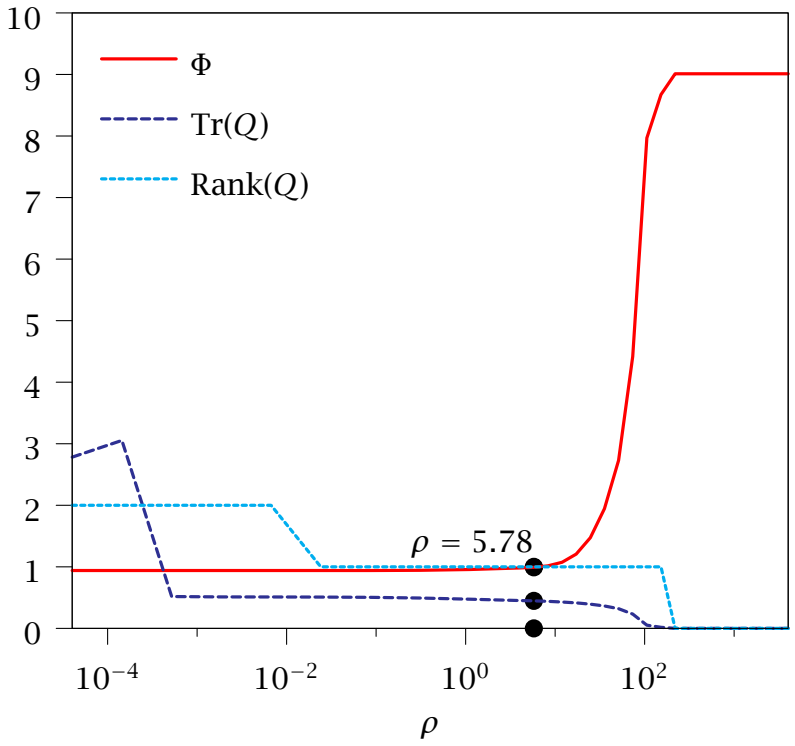


Figure 3.1: Values of competing parts of the objective function in Equation 3.23 for different values of ρ and the rank of Q

After decomposition according to Equation 3.21 we get, $\hat{G} = [0.670, 0.372]^T, \hat{Q}_w = 1$. Again if $\hat{Q}_w$ were chosen to be 0.5, then the decomposition of $\hat{Q}$ gives $\hat{G} = [0.95, 0.52]^T$, which is close to the actual G simulating the data.

The estimated positive semidefinite $\hat{Q}$ and a positive definite $\hat{R}_v$ can then be used to tune any state estimator chosen by the user. With the above estimated covariances for $\rho = 0.31$, the Kalman filter tuning $\hat{L}$ is compared with the optimal L_o:

$$\hat{L} = \begin{bmatrix} 0.312 \\ 0.211 \end{bmatrix} \qquad L_o = \begin{bmatrix} 0.328 \\ 0.202 \end{bmatrix}$$

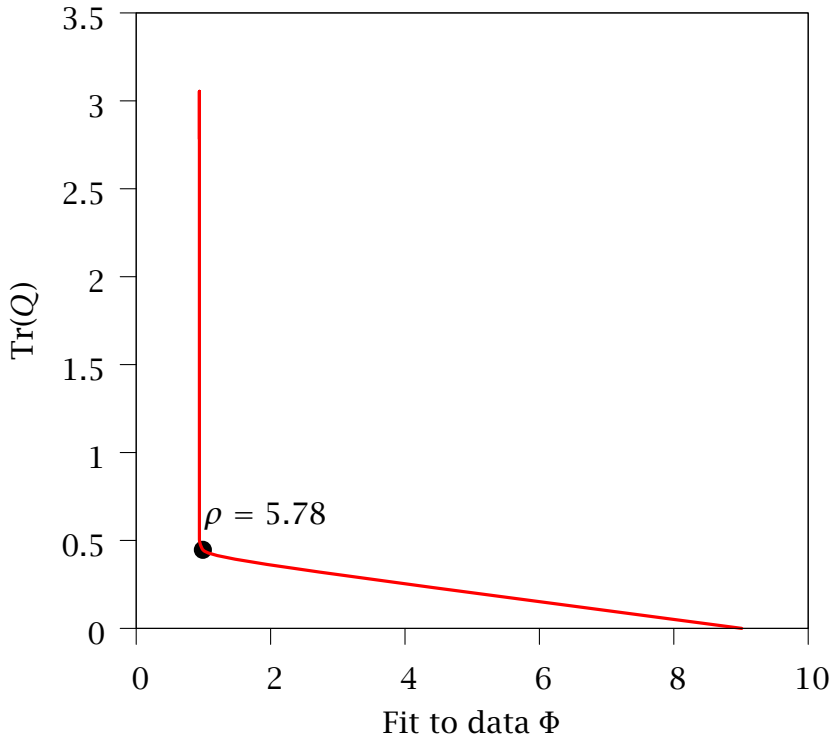


Figure 3.2: Tradeoff plot between Φ and $Tr(Q)$ from Equation 3.23 to choose the tradeoff parameter ρ

3.5 Conclusions

Given a set of system matrices A, C, a known noise shaping matrix G and an initial arbitrary stable filter gain L, uniqueness of the estimates of Q_w and R_v using the ALS technique can be checked using the simple conditions in Theorem 3.1. The computational burden in checking these conditions is minimal even for large dimension systems. One of the major uncertainties in the process industries is the disturbance structure affecting the significant variables of the plant. For linear models we showed that the disturbance structure is captured accurately by the matrix G in Equation 3.1a, which

shapes the noises entering the states. Estimation of the minimum number of disturbances affecting the states is equivalent to minimizing the rank of G. An estimation procedure using SDP and a rank heuristic was shown to give a tradeoff between fit to the data and the minimization of the rank. The 'knee' of the tradeoff curve was shown to give good estimates for the minimum number of disturbances and the disturbance structure.

3.6 Appendix

3.6.1 Proof of Lemma 3.2

Let $[q_N, r_N]^T$ be an element in the null space of $\tilde{\mathcal{A}}$ in Equation 3.20, where the dimensions are $q_N \in \mathbb{R}^{\frac{g(g+1)}{2}\times 1}$ and $r_N \in \mathbb{R}^{\frac{p(p+1)}{2}\times 1}$. This implies:

$$\tilde{\mathcal{A}} \begin{bmatrix} q_N \\ r_N \end{bmatrix} = 0 \qquad \text{or,}$$

$$\begin{bmatrix} \mathcal{A}_1(G \otimes G)\mathcal{D}_g & \mathcal{A}_2 \end{bmatrix} \begin{bmatrix} q_N \\ r_N \end{bmatrix} = 0$$

where, $\mathcal{A}_1$ and $\mathcal{A}_2$ from Equation 3.19 are:

$$\begin{aligned} \mathcal{A}_1 &= (C \otimes \mathcal{O})A^\dagger \\ \mathcal{A}_2 &= [(C \otimes \mathcal{O})A^\dagger(AL \otimes AL) + (I_p \otimes \Gamma)]\mathcal{D}_p \quad \text{and} \\ A^\dagger &= (I_{n^2} - \bar{A} \otimes \bar{A})^{-1} \end{aligned}$$

We then have:

$$(C \otimes \mathcal{O})A^\dagger(G \otimes G)\mathcal{D}_g q_N + [(C \otimes \mathcal{O})A^\dagger(AL \otimes AL) + (I_p \otimes \Gamma)]\mathcal{D}_p r_N = 0 \tag{3.25}$$

We can rewrite $\mathcal{O}$ and Γ as:

$$\mathcal{O} = \begin{bmatrix} C \\ \mathcal{O}_1\bar{A} \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} I_p \\ \mathcal{O}_1(-AL) \end{bmatrix}$$

where,

$$\mathcal{O}_1 = \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-2} \end{bmatrix}$$

If (A, C) is observable (Assumption 3.1), then $\mathcal{O}_1$ has full column rank for $N \geq (n+1)$.

Partitioning $\mathcal{O}$ and Γ as above, we can write Equation 3.25 as the following Equations:

$$(C \otimes C)A^{\dagger}[(G \otimes G)\mathcal{D}_g q_N + (AL \otimes AL)\mathcal{D}_p r_N] + \mathcal{D}_p r_N = 0 \tag{3.26a}$$

$$(C \otimes \mathcal{O}_1\bar{A})A^{\dagger}[(G \otimes G)\mathcal{D}_g q_N + (AL \otimes AL)\mathcal{D}_p r_N] + (I_p \otimes \mathcal{O}_1(-AL))\mathcal{D}_p r_N = 0 \tag{3.26b}$$

By expanding $\bar{A} = A - ALC$ and using Equation 3.26a, Equation 3.26b simplifies to:

$$(I_p \otimes \mathcal{O}_1 A)(C \otimes I_n)A^{\dagger}[(G \otimes G)\mathcal{D}_g q_N + (AL \otimes AL)\mathcal{D}_p r_N] = 0$$

Since $\mathcal{O}_1$ is full column rank (Assumption 3.1) and A is non singular (Assumption 3.3), $(I_p \otimes \mathcal{O}_1 A)$ is also full column rank. This implies:

$$(C \otimes I_n)A^{\dagger}[(G \otimes G)\mathcal{D}_g q_N + (AL \otimes AL)\mathcal{D}_p r_N] = 0 \tag{3.27}$$

Substituting Equation 3.27 in 3.26a and noting that $(C \otimes C)$ can be written as $(I_n \otimes C)(C \otimes I_n)$, we get:

$$\mathcal{D}_p r_N = 0$$

Equation 3.27 then simplifies to:

$$(C \otimes I_n)A^{\dagger}(G \otimes G)\mathcal{D}_g q_N = 0$$

Thus, q_N is an element in the null space of $M = (C \otimes I_n)(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(G \otimes G)\mathcal{D}_g$ and $r_N = 0$.

Proving the second part of the lemma is straightforward by starting with Equation 3.27 and multiplying with $(I_n \otimes \mathcal{O})$, which is full column rank. □

3.6.2 Proof of Lemma 3.3

Since the X matrix is constrained to be symmetric, we only need to consider the symmetric $p = \frac{n(n+1)}{2}$ elements of X. Let these symmetric elements of X be stacked in the vector $z \in \mathbb{R}^p$.

The original optimization in Lemma 3.3 can then be written as:

$$\begin{aligned} \min_{z} \quad & (\tilde{A}z - b)^T(\tilde{A}z - b) + d^T z \\ \text{subject to} \quad & \sum_{i=1}^{i=p} z_i B_i \geq 0 \end{aligned}$$

where, $\tilde{A}$ is the A modified to operate only on the symmetric elements of X, $B_i \in \mathbb{R}^{n \times n}, i = 1, \cdots, p$ are the basis matrices for a symmetric $\in \mathbb{R}^{n \times n}$ matrix and d is chosen appropriately such that $\rho \text{Tr}\,(X) = d^T z$.

Using the standard technique for converting quadratic objective functions into Linear Matrix Inequalities (see for example Vandenberghe and Boyd (1996, p. 54)), we

get:

$$\begin{aligned} \min_{z,t} \quad & t + d^T z \\ \text{subject to} \quad & \sum_{i=1}^{i=p} z_i B_i \succeq 0 \\ & \begin{bmatrix} t & (\tilde{A}z - b)^T \\ (\tilde{A}z - b) & I \end{bmatrix} \succeq 0 \end{aligned}$$

Define $m = p + 1$, $x = [z^T, t]^T \in \mathbb{R}^m$, $c = [d^T, 1]^T \in \mathbb{R}^m$ and the matrices

$$F_0 = \begin{bmatrix} 0 & -b^T & \\ -b & I & \\ & & 0 \end{bmatrix}, \quad F_m = \begin{bmatrix} 1 & 0 & \\ 0 & 0 & \\ & & 0 \end{bmatrix}, \quad F_i = \begin{bmatrix} 0 & a_i^T & \\ a_i & 0 & \\ & & B_i \end{bmatrix}$$

$$i = 1, \cdots, p$$

here, the 0 represents zero matrices with appropriate dimensions. We then get the final form of the optimization as a standard primal Semidefinite Programming(SDP) problem:

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & F(x) \succeq 0 \\ \text{where} \quad & F(x) \triangleq F_0 + \sum_{i=1}^{m} x_i F_i \end{aligned}$$

□

3.6.3 Some Useful Derivatives of Matrix Functions

The results below follow from Magnus and Neudecker (1999, chap. 9) and Graham (1981, chap. 5).

Given $Q \in \mathbb{R}^{n\times n}$ is a symmetric matrix and $A \in \mathbb{R}^{p\times \frac{n(n+1)}{2}}$ and $b \in \mathbb{R}^{p}$ are some arbitrary constant matrices with $p \geq n$.

$$F = (AQ_s - b)^T(AQ_s - b) + \rho \text{Tr}\,(Q) - \mu \log|Q|$$

The first and second derivatives for the above function with respect to the matrix Q are given by:

$$\left[\frac{\partial F}{\partial Q}\right]_s = 2A^TA(Q)_s - 2A^Tb + \rho(I_n)_s - \mu(Q^{-1})_s \tag{3.28}$$

$$\left[\frac{\partial^2 F}{\partial Q^2}\right]_s = 2A^TA - \mu(Q^{-1} \otimes Q^{-1}) \tag{3.29}$$

3.6.4 The ALS Technique for the Full Autocovariance Matrix

The full autocovariance matrix is defined as

$$\mathcal{R}(N) = E\begin{bmatrix} y_k y_k^T & \cdots & y_k y_{k+N-1}^T \\ \vdots & \ddots & \vdots \\ y_{k+N-1} y_k^T & \cdots & y_k y_k^T \end{bmatrix} \tag{3.30}$$

where N is the number of lags defined by the user. Using Equations 3.6, 3.7 and 3.8, we can write the autocovariance matrix as:

$$\begin{aligned} \mathcal{R}(N) = \mathcal{O}P\mathcal{O}^T + \Gamma\left[\bigoplus_{i=1}^{N} \bar{G}\bar{Q}_w\bar{G}^T\right]\Gamma^T \\ + \Psi\left[\bigoplus_{i=1}^{N} R_v\right] + \left[\bigoplus_{i=1}^{N} R_v\right]\Psi^T + \bigoplus_{i=1}^{N} R_v \end{aligned} \tag{3.31}$$

in which

$$\mathcal{O} = \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-1} \end{bmatrix} \quad \Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 \\ C & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots \\ C\bar{A}^{N-2} & \cdots & C & 0 \end{bmatrix} \quad \Psi = \Gamma \left[\bigoplus_{j=1}^{N} (-AL) \right] \tag{3.32}$$

and $\bigoplus$ is the symbol for the direct sum, satisfying:

$$\bigoplus_{i=1}^{k} Q_w = (I_k \otimes Q_w)$$

Stacking Equation 3.31 and using the stacked form of Equation 3.8 to substitute out P:

$$\begin{aligned} b = (\mathcal{R}(N))_s = & [(\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma)\mathcal{I}_{n,N}](GQ_wG^T)_s \\ & + \{[(\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma)\mathcal{I}_{n,N}](AL \otimes AL) \\ & + [\Psi \oplus \Psi + I_{p^2N^2}]\mathcal{I}_{p,N}\}(R_v)_s \end{aligned} \tag{3.33}$$

in which $\mathcal{I}_{p,N}$ is a permutation matrix to convert the direct sum to a vector, i.e. $\mathcal{I}_{p,N}$ is the $(pN)^2 \times p^2$ matrix of zeros and ones satisfying:

$$\left(\bigoplus_{i=1}^{N} R_v \right)_s = \mathcal{I}_{p,N}(R_v)_s \tag{3.34}$$

Again using Equation 3.14, we can use the ergodic property of the L-innovations to estimate the autocovariance matrix from $\mathbb{Y}$ in Equation 3.15. Using Equation 3.14, the estimate $\hat{\mathcal{R}}(N)$ is then given by:

$$\hat{\mathcal{R}}(N) = \frac{1}{N_d - N + 1} \mathbb{Y}\mathbb{Y}^T \tag{3.35}$$

and $\hat{b} = [\hat{\mathcal{R}}(N)]_s$. Given the linear relation in Equation 3.33 and an estimate of the autocovariance matrix in Equation 3.35, we can formulate the following semipositive definite constrained minimization problem:

$$\Phi = \min_{Q_w, R_v} \left\| \mathcal{A} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \hat{b} \right\|_W^2 \tag{3.36}$$
$$\text{subject to,} \quad Q_w, R_v \geq 0, \quad R_v = R_v^T, \quad Q_w = Q_w^T$$

Using Equation 3.33, we can then write $\mathcal{A}$ explicitly as:

$$\begin{aligned} \mathcal{A} &= [\mathcal{A}_1 \quad \mathcal{A}_2] \\ \mathcal{A}_1 &= [(\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma)\mathcal{I}_{n,N}](G \otimes G) \\ \mathcal{A}_2 &= [(\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma)\mathcal{I}_{n,N}](AL \otimes AL) \\ &\quad + (\Psi \oplus \Psi + I_{p^2N^2})\mathcal{I}_{p,N} \end{aligned} \tag{3.37}$$

The estimation method in Equation 3.36 is the ALS technique for the full autocovariance matrix. It can be shown by following a derivation similar to the proof to Lemma 3.2 in Appendix 3.6.1, that the conditions for uniqueness for the ALS technique with the single block column of the autocovariance matrix also hold when the full autocovariance matrix is used.

Chapter 4

Optimal Weighting of the ALS Objective and Implementation Issues [1]

This chapter lists some of the implementation issues for the ALS technique described in Chapter 3. In Section 4.1, we derive the optimal weighting matrix to calculate the linear unbiased minimum variance estimates of the covariances. The issue of w_k and v_k being correlated with each other is covered in Section 4.2. Finally numerical simplications to speed up the ALS algorithm are presented in Section 4.3.

[1] **Portions of this chapter are to appear in Rajamani and Rawlings (September, 2007)**

4.1 Minimum Variance and Optimal Weighting

Theorem 4.1. *For a linear model of the form* $y = Ax + e$ *with* $E[e] = 0$ *and* $E[ee^T] = R$, *the weighted least-squares estimator for* x *is:*

$$\min_x \|Ax - y\|^2_{R^{-1}}$$

The weighted least-squares estimator given by

$$\hat{x} = (A^T R^{-1} A)^{-1} A^T R^{-1} y$$

then has the minimum variance among all linear unbiased estimators.

This statement is a classical generalized least-squares result for the linear regression model first considered by Aitken (1935). A more recent proof can be found in Magnus and Neudecker (1999, p. 259).

The weighted least-squares estimation of the covariances is given by the ALS technique as shown by Equation 3.18 in Chapter 3 and repeated here for convenience:

$$\Phi = \min_{GQ_wG, R_v} \left\| \mathcal{A} \begin{bmatrix} \mathcal{D}_n(GQ_wG^T)_{ss} \\ (R_v)_{ss} \end{bmatrix} - \hat{b} \right\|^2_W \tag{4.1}$$

$$\text{subject to,} \quad GQ_wG^T, R_v \geq 0, \quad R_v = R_v^T$$

In Odelson et al. (2006b) and in Chapter 3, the weighting matrix W in the ALS problem is taken to be the identity matrix. The minimum variance property for the estimates then does not hold. We next derive the formula for the minimum variance weighting matrix W.

Following the analogy of Theorem 4.1 for Equation 4.1, if $\hat{b}$ is an unbiased estimator of b, then $b = E[\hat{b}]$. Define $S \triangleq E[(\hat{b} - b)(\hat{b} - b)^T] = \text{cov}(\hat{b})$ as the covariance of $\hat{b}$. Then $W = S^{-1}$ is the weighting that gives minimum variance for the ALS problem. It is shown in Odelson et al. (2006b) that $\hat{b}$ in Equation 4.1 and calculated using Equation 3.16 in Chapter 3 is an unbiased estimator.

Lemma 4.1. *Given the L-innovations from Equation 3.4 and the definition of* $\mathbb{Y}$ *from Equation 3.15 in Chapter 3, we have:*

$$E[\mathbb{Y}] = 0$$

$$E[\mathbb{Y}\mathbb{Y}^T] \triangleq E[\mathbb{Y}_s\mathbb{Y}_s^T] = \Omega$$

with Ω *as defined in Appendix 4.4.1 (Equation 4.25). The random matrix* $\mathbb{Y}$ *is distributed normally with* $\mathbb{Y} \sim N(0, \Omega)$.

Proof of Lemma 4.1 is given in Appendix 4.4.1.

Note that the formula for Ω as given by Equation 4.25 depends on the unknown disturbance covariances Q_w, R_v and G.

Theorem 4.2. *The minimum variance weight to use in the ALS objective in Equation 4.1 is given by* $W = S^{\dagger}$, *where,*

$$S = \frac{T(I_{\tilde{n}^2\tilde{p}^2} + K_{(\tilde{n}\tilde{p})(\tilde{n}\tilde{p})})((K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}) \otimes (K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}))T^T}{(N_d - N + 1)^2} \tag{4.2}$$

and K_{ij} *is the commutation matrix defined in Magnus and Neudecker (1979).* T *is defined as:*

$$T = (\mathbb{E} \otimes I_p)(I_{\tilde{p}^2} \otimes (I_{\tilde{n}})_s)^T(I_{\tilde{p}} \otimes K_{\tilde{p}\tilde{n}} \otimes I_{\tilde{n}})$$

and $\mathbb{E} = [I_p, 0 \cdots 0]$. $\tilde{n} = N_d - N + 1$ *and* $\tilde{p} = Np$

Proof. $\mathbb{Y} \in \mathbb{R}^{\tilde{p} \times \tilde{n}}$ is the matrix given in Equation 3.15 and is normally distributed with mean 0 and covariance Ω as defined in Lemma 4.1. The fourth moment of $\mathbb{Y}$ is defined as follows:

$$\text{cov}\,[\mathbb{Y}\mathbb{Y}^T] \triangleq \text{cov}((\mathbb{Y}\mathbb{Y}^T)_s)$$

The formula for the fourth moment of $\mathbb{Y}$ i.e $\text{cov}(\mathbb{Y}\mathbb{Y}^T)$ for a normal distribution is given by:

$$\text{cov}\,(\mathbb{Y}\mathbb{Y}^T) = T_1(I_{\tilde{n}^2\tilde{p}^2} + K_{(\tilde{n}\tilde{p})(\tilde{n}\tilde{p})})((K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}) \otimes (K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}))T_1^T \tag{4.3}$$

where, $T_1 = (I_{\tilde{p}^2} \otimes (I_{\tilde{n}})_s)^T(I_{\tilde{p}} \otimes K_{\tilde{p}\tilde{n}} \otimes I_{\tilde{n}})$. The formula follows from the results in Ghazal and Neudecker (2000). See Ghazal and Neudecker (2000) for more details on the derivation. The commutation matrix K_{ij} is a $\in \mathbb{R}^{ij \times ij}$ matrix containing only 1's and 0's and gives the following relationship between $(A)_s$ and $(A^T)_s$ when A is a $\in \mathbb{R}^{i \times j}$ matrix: $(A)_s = K_{ij}(A^T)_s$ and $(A^T)_s = K_{ji}(A)_s$.

We also have from Equations 3.16 and 3.17 in Chapter 3:

$$\begin{aligned}\hat{b} &= (\mathbb{Y}\mathbb{Y}^T\mathbb{E}^T)_s \\ &= (\mathbb{E} \otimes I_p)(\mathbb{Y}\mathbb{Y}^T)_s\end{aligned}$$

From Equation 4.3 we can then calculate the covariance of $\hat{b}$:

$$\begin{aligned}S &= \text{cov}(\hat{b}) \\ &= \frac{(\mathbb{E} \otimes I_p)\text{cov}(\mathbb{Y}\mathbb{Y}^T)(\mathbb{E}^T \otimes I_p)}{(N_d - N + 1)^2}\end{aligned}$$

Thus, we get Equation 4.2 as the covariance of $\hat{b}$ with $T = (\mathbb{E} \otimes I_p)T_1$.

The optimal weight is then $W = S^{-1}$ following Theorem 4.1. If S is singular, then without loss of generality we can take $W = S^{\dagger}$, the Moore-Penrose pseudo-inverse of S. □

The weight W is a complicated function depending on the values of the unknown covariances. A recursive calculation may be carried out for calculating W and the covariances.

1. Guess a value for $\hat{Q}, \hat{R_v}$, where $Q = GQ_wG^T$ and calculate Ω and $W = S^{-1}$ using Equations 4.25 and 4.2.

2. Use the estimated weight in the ALS technique to estimate $\hat{Q}, \hat{R_v}$ using Equation 4.1

3. Use estimates from the previous step to recalculate W

4. Iterate until convergence

The convergence of the above iterative scheme has not been tested because of the significant computational burden (see Remark 4.2).

Remark 4.1. *If the initial estimator gain L is optimal, the L-innovations (which are the optimal innovations) are white. The formula for S (Equation 4.2) is then simpler than Equation 4.2 and is the second moment of the Wishart distribution (Anderson, 2003, p. 255). White innovations also implies optimality of the filter and there would be no need*

to calculate the covariances. In the more practical situation when the L-innovations are not white, the assumption of 'whiteness' leads to an incorrect weighting. This incorrect weighting was used in Dee et al. (1985).

Remark 4.2. *The computation of S from Equation 4.2 becomes prohibitively large even for a small dimensional problem with large data sets. This is a drawback for any practical application until efficient means for the computation are developed.*

Remark 4.3. *Although the weight may be estimated from data, a large data set is required before getting reliable estimates for the weights. An attractive alternative to circumvent the need for large data sets is to use Bootstrapping (for eg. Stoffer and Wall (1991)).*

4.1.1 Example of Lower Variance

Consider the following model for the system:

$$x_{k+1} = \begin{bmatrix} 0.732 & -0.086 \\ 0.172 & 0.990 \end{bmatrix} x_k + \begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix} w_k$$

$$y_k = x_k + v_k$$

Data is generated by drawing the noises from the following distributions:

$$w_k \sim N\left(0, \begin{bmatrix} 0.5 & 0 \\ 0 & 0.2 \end{bmatrix}\right), \qquad v_k \sim N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

The ALS estimation of the covariances Q_w, R_v for a set of data simulated using $W = I$ and using the minimum variance weight (iterative scheme) from the above section is

compared. The covariance estimation is repeated 100 times and the results are plotted to check for the variance in the estimates. The diagonal elements of the estimated Q_w, R_v are plotted.

As seen in Figures 4.1 and 4.2 using the optimal weight gives estimates having much lower variance than using $W = I$.

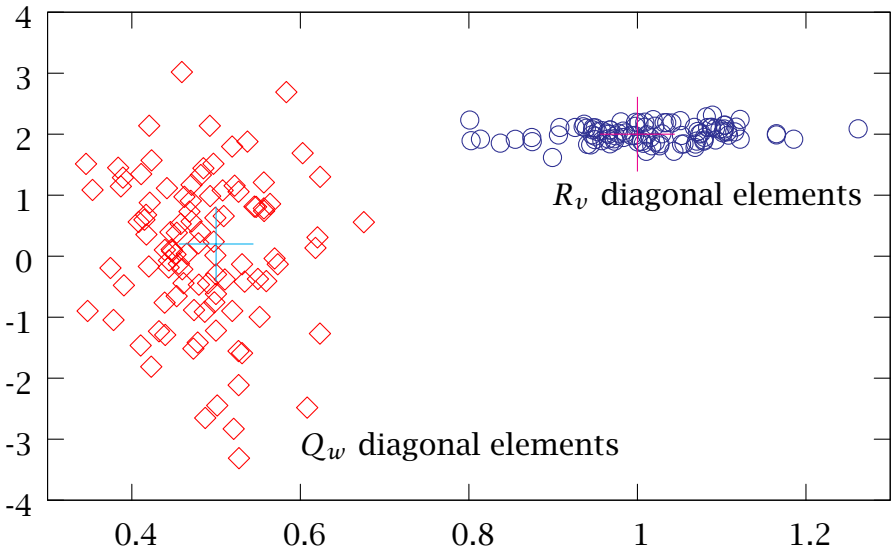


Figure 4.1: Covariance estimates using $W = I$ in ALS

4.1.2 Conclusions

Estimates of the noise covariances from data are minimum variance only if the least-squares is weighted with the optimal weight. This weight was shown to depend on the fourth moment of data and a formula was derived (Theorem 4.2). An example was presented to show the reduced variance in the covariance estimates when using the minimum variance weight. The complicated nature of the formulae do not make them

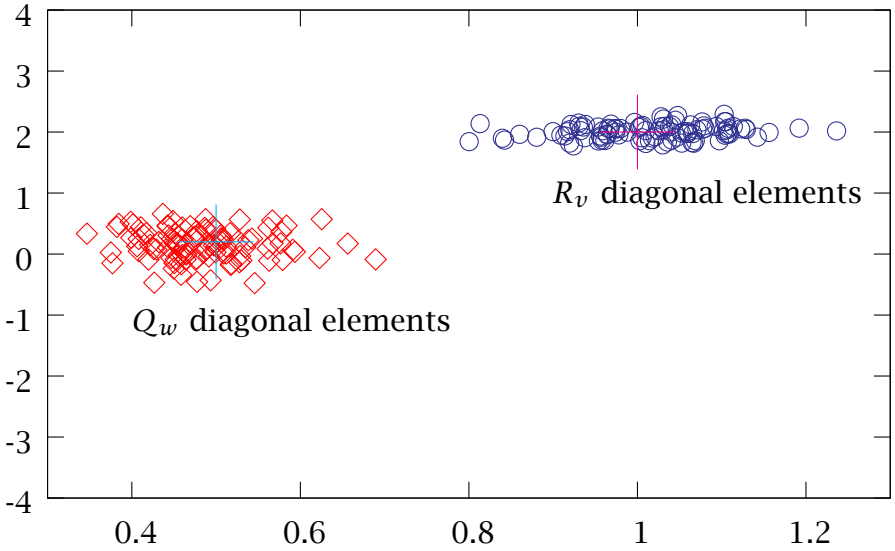


Figure 4.2: Covariance estimates using the minimum variance weight in ALS

practical using current computational techniques. The weight however puts to rest the issue of the existence of the best linear unbiased estimator for the covariances.

4.2 Estimating the Cross-Covariance Term from Data

We start with the linear state-space model:

$$x_{k+1} = Ax_k + Bu_k + Gw_k \tag{4.4a}$$

$$y_k = Cx_k + v_k \tag{4.4b}$$

Often in physical systems, the noises w_k, v_k are independent. If however w_k, v_k are not independent (for example, when part of the noise in the inputs feeds into the measurements), then there is an associated nonzero cross-covariance given by $S = E[w_k v_k^T]$.

Writing the state-space model in terms of estimate errors $\varepsilon_k = x_k - \hat{x}_k$ starting with an initial arbitrary stable filter gain L as in Section 3.1 in Chapter 3, we get:

$$\varepsilon_{k+1} = \underbrace{(A - ALC)}_{\bar{A}} \varepsilon_k + \underbrace{\begin{bmatrix} G & -AL \end{bmatrix}}_{\bar{G}} \underbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}_{\bar{w}_k}$$

$$y_k = C\varepsilon_k + v_k$$

The noise covariances are given by:

$$E[\bar{w}_k \bar{w}_k^T] = \bar{Q}_w = \begin{bmatrix} Q_w & S \\ S^T & R_v \end{bmatrix} \text{ and } \qquad E[\bar{w}_k v_k^T] \begin{bmatrix} S \\ R_v \end{bmatrix}$$

The autocovariances of the L-innovations in this case are given by:

$$E[y_k y_k^T] = CPC^T + R_v \tag{4.5}$$

$$E[y_{k+j} y_k^T] = C\bar{A}^j PC^T - C\bar{A}^{j-1} ALR_v + C\bar{A}^{j-1} GS \tag{4.6}$$

and P is given by the Lyapunov equation:

$$P = \bar{A} P \bar{A}^T + GQ_w G^T + ALR_v L^T A^T - ALSG^T - GSL^T A^T \tag{4.7}$$

Repeating the analysis as in Appendix 3.6.4 of Chapter 3, we define the full autocovariance matrix as in Equation from 3.30:

$$\mathcal{R}(N) = E \begin{bmatrix} y_k y_k^T & \cdots & y_k y_{k+N-1}^T \\ \vdots & \ddots & \vdots \\ y_{k+N-1} y_k^T & \cdots & y_k y_k^T \end{bmatrix} \tag{4.8}$$

Using Equations 4.5, 4.6 and 4.7, we then get:

$$\mathcal{R}(N) = \begin{bmatrix} CPC^T + R_v & \cdots & CP(\bar{A}^T)^{N-1}C^T - R_v L^T A^T (\bar{A}^T)^{N-2} C^T + S^T G^T (\bar{A}^T)^{N-2} C^T \\ \vdots & \ddots & \vdots \\ C\bar{A}^{N-1}PC^T - C\bar{A}^{N-2}ALR_v + C\bar{A}^{N-2}GS & \dots & CPC^T + R_v \end{bmatrix} \tag{4.9}$$

In a more compact notation, the above equation can be written as (substituting Equation 4.7):

$$\begin{aligned} \mathcal{R}(N) = \mathcal{O}P\mathcal{O}^T + \Gamma \left[\bigoplus_{i=1}^{N} \bar{G}\bar{Q}_w\bar{G}^T \right] \Gamma^T + \Psi \left[\bigoplus_{i=1}^{N} R_v \right] + \left[\bigoplus_{i=1}^{N} R_v \right] \Psi^T + \bigoplus_{i=1}^{N} R_v \\ + \mathbb{Z} \left[\bigoplus_{i=1}^{N} S \right] + \left[\bigoplus_{i=1}^{N} S \right] \mathbb{Z}^T \end{aligned} \tag{4.10}$$

in which

$$\mathcal{O} = \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-1} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 \\ C & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots \\ C\bar{A}^{N-2} & \cdots & C & 0 \end{bmatrix}, \quad \Psi = \Gamma \left[\bigoplus_{j=1}^{N} (-AL) \right], \quad \mathbb{Z} = \Gamma \left[\bigoplus_{i=1}^{N} G \right] \tag{4.11}$$

Note the appearance of the extra terms involving $\mathbb{Z}$ and S compared to Equation 3.31 when there was no cross-covariance. The column-wise stacking operation performed on Equations 4.7 and 4.11 gives:

$$P_s = (I_n^2 - \bar{A} \otimes \bar{A})^{-1}[(G \otimes G)(Q_w)_s + (AL \otimes AL)(R_v)_s - (G \otimes AL)S_s - (AL \otimes G)S_s$$

and

$$
\begin{aligned}
[\mathcal{R}(N)]_s = {} & [(\mathcal{O}\otimes\mathcal{O})(I_n^2 - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N}](G\otimes G)(Q_w)_s \\
& + \left\{[(\mathcal{O}\otimes\mathcal{O})(I_n^2 - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N}](AL\otimes AL) + [\Psi\oplus\Psi + I_{p^2N^2}]\mathcal{I}_{p,N}\right\}(R_v)_s \\
& + \left\{[(\mathcal{O}\otimes\mathcal{O})(I_n^2 - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N}][(-AL\otimes G) - (G\otimes AL)K_{gp}]\right. \\
& \left. + [\mathbb{Z}\oplus\mathbb{Z}]\mathcal{I}_{n,p,N}\right\}(S)_s
\end{aligned}
\tag{4.12}
$$

where we define $\mathcal{I}_{n,p,N}$ as the $\mathbb{R}^{npN^2\times np}$permutation matrix containing 1's and 0's giving the relation:

$$
\left(\bigoplus_{i=1}^{N} S\right)_s = \mathcal{I}_{n,p,N}(S)_s
$$

and K_{ij} is the commutation matrix as given in Section 4.1. Thus, if $b = [\mathcal{R}(N)]_s$, then the full matrix ALS optimization with the cross-covariance term included can be formulated as (with $W = I$):

$$
\min_{Q_w,R_v,S} \left\| \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 & \mathcal{A}_3 \end{bmatrix} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \\ (S)_s \end{bmatrix} \right\|^2 \tag{4.13}
$$

$$
\text{subject to,} \quad Q_w \geq 0, R_v \geq 0
$$

with the following definitions for $\mathcal{A}$:

$$
\mathcal{A}_1 = [(\mathcal{O}\otimes\mathcal{O})(I_n^2 - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N}](G\otimes G)
$$

$$
\mathcal{A}_2 = [(\mathcal{O}\otimes\mathcal{O})(I_n^2 - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N}](AL\otimes AL) + [\Psi\oplus\Psi + I_{p^2N^2}]\mathcal{I}_{p,N}
$$

$$
\mathcal{A}_3 = [(\mathcal{O}\otimes\mathcal{O})(I_n^2 - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N}][(-AL\otimes G) - (G\otimes AL)K_{gp}] + [\mathbb{Z}\oplus\mathbb{Z}]\mathcal{I}_{n,p,N}
$$

Åkesson et al. (2007) present a variation of the above derivation to include integrating disturbance models. Åkesson et al. (2007) also present an interior-point predictor-corrector algorithm to solve the above SDP formulation with the cross-term.

4.2.1 ALS Column Formulation for Estimating the Cross-Covariance

We next repeat the above derivation using only the first column block of the autocovariance matrix (ACM). Using the full autocovariance matrix has the advantage of giving a marginally better conditioning of the $\mathcal{A}$ matrix in the ALS optimization. The symmetry constraints on Q_w, R_v are implicit when the problem is formulated with the full autocovariance matrix. On the other hand, using only the first column block of the ACM has the advantage of a reduced computational burden as opposed to using the full ACM.

If only the first block column of the ACM were used, we get:

$$\mathcal{R}_1(N) = E\begin{bmatrix} y_k y_k^T \\ y_{k+1} y_k^T \\ \vdots \\ y_{k+N-1} y_k^T \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-1} \end{bmatrix}}_{\mathcal{O}} PC^T + \underbrace{\begin{bmatrix} I_p \\ -CAL \\ \vdots \\ -C\bar{A}^{N-2}AL \end{bmatrix}}_{\Gamma_1} R_v + \underbrace{\begin{bmatrix} 0 \\ CG \\ \vdots \\ C\bar{A}^{N-2}G \end{bmatrix}}_{\mathbb{Z}_1} S$$

Column-wise stacking of the above Equation gives:

$$\begin{aligned}[\mathcal{R}_1(N)]_s &= (C \otimes \mathcal{O})(I_n^2 - \bar{A} \otimes \bar{A})^{-1}(G \otimes G)(Q_w)_s \\ &+ [(C \otimes \mathcal{O})(I_n^2 - \bar{A} \otimes \bar{A})^{-1}(AL \otimes AL) + (I_p \otimes \Gamma_1)](R_v)_s \\ &+ \left\{(C \otimes \mathcal{O})(I_n^2 - \bar{A} \otimes \bar{A})^{-1}[(-AL \otimes G) - (G \otimes AL)K_{gp}] + (I_p \otimes \mathbb{Z}_1)\right\}(S)_s\end{aligned}$$

Thus if $b_1 = [\mathcal{R}_1(N)]_s$, then the single block column ALS with the cross-covariance term included can be formulated as (with $W = I$):

$$\min_{Q_w, R_v, S} \left\| \begin{bmatrix} \tilde{\mathcal{A}}_1 & \tilde{\mathcal{A}}_2 & \tilde{\mathcal{A}}_3 \end{bmatrix} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \\ (S)_s \end{bmatrix} \right\|^2 \tag{4.14}$$

$$\text{subject to,} \quad Q_w \geq 0, R_v \geq 0$$

and the modified matrices are:

$$\tilde{\mathcal{A}}_1 = (C \otimes \mathcal{O})(I_n^2 - \bar{A} \otimes \bar{A})^{-1}(G \otimes G)$$

$$\tilde{\mathcal{A}}_2 = (C \otimes \mathcal{O})(I_n^2 - \bar{A} \otimes \bar{A})^{-1}(AL \otimes AL) + (I_p \otimes \Gamma_1)$$

$$\tilde{\mathcal{A}}_3 = (C \otimes \mathcal{O})(I_n^2 - \bar{A} \otimes \bar{A})^{-1}[(-AL \otimes G) - (G \otimes AL)K_{gp}] + (I_p \otimes \mathbb{Z}_1)$$

4.2.2 Alternative Techniques for Estimating the Cross-Covariance

If L_o is the optimal Kalman filter for the linear state-space model in Equations 4.4a and 4.4b with w_k, v_k having cross correlations, the innovations form of the model can be written as:

$$\begin{aligned} x_{k+1} &= \bar{A}x_k + AL_o e_k \\ y_k &= Cx_k + e_k \end{aligned} \tag{4.15}$$

Here e_k are the optimal innovations. If the above equation is the starting point for our estimation procedure, we can modify the ALS optimization as follows.

The state noise covariance in the innovations form is $\text{cov}(AL_o e_k) = AL_o S L_o^T A^T$, the measurement noise covariance is $\text{cov}(e_k) = S$ and the cross-covariance is $\text{cov}(AL_o e_k, e_k) =$

AL_oS. Modifying the original ALS optimization from Equation 4.1 to apply to the model in Equation 4.15 we get:

$$\min_{L_o,S} \left\| \mathcal{A} \begin{bmatrix} (AL_oSL_o^T)_s \\ S_s \\ (AL_oS)_s \end{bmatrix} - b \right\|^2$$

$$= \min_{L_o,S} \left\| \mathcal{A} \begin{bmatrix} (AL_o \otimes AL_o) \\ I_{p^2} \\ (I_p \otimes AL_o \end{bmatrix} S_s - b \right\|^2$$

The above minimization is nonlinear in the unknowns L_o, S and there is no technique to avoid the added complexity. We thus prefer starting with the more general model in Equations 4.4a, 4.4b and minimizing a convex quadratic objective as in Section 4.2.1 instead.

Another formulation involves defining the new stochastic variable $\bar{w}_k = w_k - SR_v^{-1}v_k$ (Anderson and Moore, 1979, p.115). We then have:

$$x_{k+1} = Ax_k + Bu_k + G\bar{w}_k + GSR_v^{-1}v_k$$

$$\implies x_{k+1} = (A - GSR_v^{-1}C)x_k + \begin{bmatrix} B & GSR_v^{-1} \end{bmatrix} \begin{bmatrix} u_k \\ y_k \end{bmatrix} + G\bar{w}_k, \quad \text{and}$$

$$y_k = Cx_k + v_k$$

The inputs to the state x_{k+1} in the modified problem is now $[u_k^T, y_k^T]^T$. The new noises

$\bar{w}_k$ and v_k are not cross correlated. We have:

$$E\left\{\begin{bmatrix}\bar{w}_k \\ v_k\end{bmatrix}\begin{bmatrix}w_k^T & v_k^T\end{bmatrix}\right\} = \begin{bmatrix}Q_w - SR_v^{-1}S^T & 0 \\ 0 & R_v\end{bmatrix}$$

However, we need prior information about the measurement noise covariance R_v to formulate the problem in this manner. An alternative procedure would be recursive substitution of the estimated $\hat{R}_v$ starting with an initial guess. Again the recursive procedure makes the problem similar to a nonlinear estimation problem as opposed to the convex ALS technique in Section 4.2.1.

4.2.3 Conditions for Uniqueness with Cross-Covariance term

The uniqueness of the ALS optimization with the cross-covariance term has no known simple form as in the case when there are no cross correlations. The covariance estimates are unique if the matrix $\mathcal{A} = \begin{bmatrix}\mathcal{A}_1 & \mathcal{A}_2 & \mathcal{A}_3\end{bmatrix}$ for the full autocovariance matrix formulation of the ALS has full column rank. Similar conditions can be written when using only the first block of the autocovariance matrix.

4.3 Mathematical Simplications to Speed up the Computation of the ALS Technique

We present here mathematical simplications that can be included in the ALS code to obtain efficient and fast solutions to the optimization.

1. Using Kronecker product to calculate the permutation matrix $\mathcal{I}_{n,N}$

2. Simplifications for the Inverse

3. Estimating only the diagonal elements of the covariances

4. Using the single block column formulation of the ALS technique

5. Using efficient interior-point algorithms to solve the semidefinite optimization

4.3.1 Simplification using Kronecker products

The permutation matrix $\mathcal{I}_{i,j}$ appears in Chapter 3 in Equation 3.34. This permutation matrix is usually difficult to code and implement and requires large storage memory even for small dimension problems. For example, when the window size is $N = 15$, number of states is $n = 25$ and number of measurements are $p = 10$, the permutation matrix for implementing the ALS technique in its full autocovariance matrix formulation has the following dimensions:

$$\mathcal{I}_{N,n} \in \mathbb{R}^{(Nn)^2 \times n^2} = \mathbb{R}^{140625 \times 225}$$

The $I_{N,n}$ matrix then has 3.1×10^7 elements which are either 1 or 0. To simplify the coding of the permutation matrix, we use the following properties of Kronecker products:

$$\begin{aligned} \left(\Gamma \left[\bigoplus_{i=1}^{N} (GQ_wG^T) \right] \Gamma^T \right)_s &= (\Gamma \otimes \Gamma) \left[\bigoplus_{i=1}^{N} (GQ_wG^T) \right]_s \\ &= (\Gamma \otimes \Gamma)\mathcal{I}_{n,N}(G \otimes G)(Q_w)_s \end{aligned} \tag{4.16}$$

in which, $\mathcal{I}_{n,N}$ is a $(nN)^2 \times n^2$ matrix satisfying:

$$\left(\bigoplus_{i=1}^{N} GQ_wG^T\right)_s = \mathcal{I}_{n,N}(GQ_wG^T)_s$$

An alternate, computationally less expensive expression is given as (let $Q = GQ_wG^T$ for notational simplification):

$$\begin{aligned}\left(\Gamma\left[\bigoplus_{i=1}^{N} Q\right]\Gamma^T\right)_s &= (\Gamma(I_N \otimes Q)\Gamma)_s \\ &= (\Gamma \otimes \Gamma)(I_N \otimes Q)_s \\ &= (\Gamma \otimes \Gamma)\left[(I_N \otimes K_{nN})((I_N)_s \otimes I_n) \otimes I_n\right](Q)_s\end{aligned} \tag{4.17}$$

In the above expression K_{nN} is called the commutation matrix (Magnus and Neudecker, 1979) (implemented in Octave as *commutation_matrix(n,N)*). The third equality in the above expression uses the formula from Magnus and Neudecker (1999, p.48).

Comparing Equations 4.16 and 4.17, an alternative expression for calculating $\mathcal{I}_{n,N}$ is:

$$\mathcal{I}_{n,N} = \left[(I_N \otimes K_{nN})((I_N)_s \otimes I_n) \otimes I_n\right] \tag{4.18}$$

If $e_i, i = 1 \cdots N$ is the ith column of the I_N matrix, then $I_N = \sum_{i=1}^{N} e_i e_i^T$ and $(I_N)_s = \sum_{i=1}^{N}(e_i \otimes e_i)$.

$$\begin{aligned}\mathcal{I}_{n,N} &= (I_N \otimes K_{nN})((I_N)_s \otimes I_n) \otimes I_n \\ &= (I_N \otimes K_{nN})(\sum_{i=1}^{N}(e_i \otimes e_i) \otimes I_n) \otimes I_n \\ &= \sum_{i=1}^{N}(e_i \otimes K_{nN}(e_k \otimes I_n)) \otimes I_n \\ &= \sum_{i=1}^{N}(e_i \otimes I_n \otimes e_i \otimes I_n)\end{aligned} \tag{4.19}$$

In the above simplification we use the following properties of Kronecker products and of the commutation matrix (Magnus and Neudecker, 1999, chap. 2):

- If $A(m \times n)$ and $B(p \times q)$ are two matrices, then, $K_{pm}(A \otimes B) = (B \otimes A)$.
- $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.
- $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ if the matrices are dimensionally conforming for multiplication.

Thus, we have,

$$\begin{aligned}(\Gamma \otimes \Gamma)\mathcal{I}_{n,N} &= (\Gamma \otimes \Gamma)\left[\sum_{i=1}^{N}(e_i \otimes I_n \otimes e_i \otimes I_n)\right] \\ &= \sum_{i=1}^{N}(\Gamma(e_i \otimes I_n) \otimes \Gamma(e_i \otimes I_n))\end{aligned} \tag{4.20}$$

The expression in Equation 4.20 avoids the expensive calculation of the $\mathcal{I}_{n,N}$ matrix and the more expensive multiplication of the $(\Gamma \otimes \Gamma)$ with $\mathcal{I}_{n,N}$.

4.3.2 Simplifications for the Inverse

When implementing either the single block column or the full autocovariance matrix version of the ALS technique, there is need to compute the matrix $A^{\dagger}$ defined as:

$$A^{\dagger} = (I_{n^2} - (A - ALC) \otimes (A - ALC))^{-1} \tag{4.21}$$

$A^{\dagger}$ is a square matrix of dimension equal to the square of the state dimension. It is not uncommon in industrial applications to have number of states in the order of $50 - 100$.

This state dimension is further increased when integrating disturbances are augmented to the model states to remove steady state offset in the closed-loop.

Taking the inverse of this large dimensioned square matrix can take up significant computation time. We can however, take advantage of the structure given by the Kronecker products in $A^{\dagger}$ to avoid taking the inverses.

Let m_i be the ith column of $A^{\dagger}$ with dimensions $\mathbb{R}^{n^2 \times 1}$. From the above Equation 4.21 we have:

$$m_i = (I_{n^2} - (A - ALC) \otimes (A - ALC))^{-1} e_i$$

where, e_i is the ith column of the identity matrix, I_{n^2}. Let the elements of m_i be stacked column-wise in a $\mathbb{R}^{n \times n}$ matrix M_i, and the elements of e_i stacked column-wise in the matrix E_i. Following rules of Kronecker products, we then have:

$$M_i - (A - ALC) M_i (A - ALC)^T = E_i$$

M_i thus is the solution to the Lyapunov equation. Efficient algorithms for finding the solution to the Lyapunov equation have been developed and included in standard software packages like Octave and Matlab (Kitagawa, 1977; Hammarling, 1982; Barnett, 1974).

The inverse $A^{\dagger}$ can thus be found using the following steps:

Step 1: Extract the ith column of the n^2 dimensioned identity matrix and stack its elements column-wise into the matrix E_i.

Step 2: Solve the discrete Lyapunov equation for M_i with E_i and $(A - ALC)$ as the arguments. The solution M_i exists and is unique since $(A - ALC)$ is stable and E_i is a

semidefinite matrix.

Step 3: Stack the elements of the matrix M_i into the $n^2 \times 1$ vector m_i. The stacked vector m_i is then the ith column of the $A^\dagger$.

Often, when implementing the ALS technique the matrix $A^\dagger$ appears as the product $A^\dagger(G \otimes G)$ or $A^\dagger(AL \otimes AL)$. The above simplification can be extended to evaluate these products using the same steps as listed, only changing Step 2 and solving the Lyapunov equation with GE_iG^T or $ALE_iL^TA^T$ as arguments instead of E_i.

4.3.3 Other Simplifications

The following simplifications can be further made to the ALS code to increase the speed of computation when used on large dimension problems.

In Chapter 3, the single block column and the full matrix formulation for the ALS was presented in Section 3.2 and Appendix 3.6.4. In Odelson (2003) it is mentioned that the full matrix formulation gives a better conditioned least-squares objective for the ALS than the single block column formulation. In addition, in the full matrix formulation the symmetry constraints on the covariance estimates are automatically imposed.

We prefer, however, the single block column formulation for the ALS technique. In all our examples tested, the single column formulation for the ALS gave a similar conditioned objective as the full matrix formulation. The computation time was significantly lower when using the single column. In addition, as seen in Chapter 3 imposing the symmetry constraints in the ALS is easily achieved through the use of the *duplication*

matrix $\mathcal{D}_n$ (Magnus and Neudecker, 1999, p. 49). We therefore, recommend using the single column formulation for the ALS with the *duplication matrix* for practical implementations due to the ease of computation of the $\mathcal{A}$ matrix in the ALS objective given in Equation 3.19.

The choice of a diagonal matrix for Q_w, R_v is often made in the industry for the ease of choosing fewer parameters. The ALS technique is easily extended to estimate only the diagonal elements of the covariances by modifying the objective and purging the columns of the $\mathcal{A}$ matrix that correspond to the non-diagonal elements. For example if $n = 2, p = 1$ with the elements in Q_w given by $Q_w = \begin{bmatrix} q_1 & q_2 \\ q_2 & q_3 \end{bmatrix}$, the ALS objective in Equation 4.1 can be written as:

$$\left\| \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 & \mathcal{A}_3 & \mathcal{A}_4 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ R_v \end{bmatrix} - \hat{b} \right\|^2$$

When the ALS technique is modified to estimate only the diagonal elements the ALS objective above is rewritten as:

$$\left\| \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_3 & \mathcal{A}_4 \end{bmatrix} \begin{bmatrix} q_1 \\ q_3 \\ R_v \end{bmatrix} - \hat{b} \right\|^2$$

Since, the objective in the ALS formulation is convex and the semidefinite constraints are also convex, simple Newton-like algorithms can be used to solve the opti-

mization. More complex algorithms that are specific to semidefinite programming can also be applied to improve the speed and efficiency of the algorithms. Åkesson et al. (2007) present a new predictor-corrector algorithm for solving the ALS optimization also including the cross-terms for the covariances. Many efficient algorithms for solving semidefinite programming problems and optimization problems involving linear matrix inequalities (LMIs) are available in the literature. The following references present some recent progress in the field of semidefinite optimization: Boyd et al. (1994); Wolkowicz et al. (2000); Vandenberghe and Boyd (1995); VanAntwerp and Braatz (2000) and Vandenberghe and Boyd (1996).

4.4 Appendix

4.4.1 Proof of Lemma 4.1

Let $Y_p = [y_1^T \cdots y_{N_d}^T]^T$. Then we have,

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \\ \hdashline y_2 \\ \vdots \\ y_{N+1} \\ \hdashline \vdots \\ \hdashline y_{N_d-N+1} \\ \vdots \\ y_{N_d} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \cdots & 0 & \cdots & & \\ \vdots & \ddots & \vdots & \vdots & \cdots & \cdots \\ 0 & \cdots & 1 & \cdots & & \\ \vdots & 1 & \cdots & 0 & & \\ 0 & \vdots & \ddots & \vdots & \cdots & \cdots \\ \vdots & 0 & \cdots & 1 & \ddots & \\ & & \cdots & \cdots & \cdots & \ddots \\ & & \cdots & \cdots & \cdots & I_N \end{bmatrix}}_{\mathbb{E}_1} \begin{bmatrix} y_1 \\ \vdots \\ y_N \\ \hdashline y_{N+1} \\ \vdots \\ y_{N_d} \end{bmatrix} \tag{4.22}$$

Thus, $\mathbb{Y}_s = \mathbb{E}_1 Y_p$. Now we find the distribution of Y_p.

From Equation 3.4, we have

$$\varepsilon_k = \bar{A}^k \varepsilon_0 + \sum_{j=0}^{k-1} \bar{A}^{k-j-1} \bar{G} \begin{bmatrix} w_j \\ v_j \end{bmatrix} \tag{4.23}$$

Taking the expectation of the above expression and noting that $E[v_k] = E[w_k] = 0$, we get,

$$E[\varepsilon_k] = \bar{A}^k E[\varepsilon_0] = 0$$

The equality follows from the stability of the initial filter gain L since for k large enough, we have $\bar{A}^k = (A - ALC)^k \approx 0$.

Taking the expectation of the L-innovations in Equation 3.4, we get:

$$E[y_j] = CE[\varepsilon_j] + E[v_k] = 0$$

holding for all $j \geq k$ (k is the initial period of transience, when for $i < k$, $E[\varepsilon_i]$ cannot be approximated as 0). Thus, we have

$$E[Y_p] = E\begin{bmatrix} y_1 \\ \vdots \\ y_{N_d} \end{bmatrix} = 0$$

Since $E(Y_p) = 0$, the covariance of Y_p is also its the second moment.

Now, calculate Ω_p the second moment of Y_p as follows:

$$\Omega_p = E\left[\begin{pmatrix} y_1 \\ \vdots \\ y_{N_d} \end{pmatrix}\begin{pmatrix} y_1^T & \cdots & y_{N_d}^T \end{pmatrix}\right]$$

Using Equations 3.6, 3.7 and 3.8 in Chapter 3, we get:

$$\begin{aligned}\Omega_p = &\underbrace{\begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N_d-1} \end{bmatrix}}_{\mathcal{O}} P\mathcal{O}^T + \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} + \Psi \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} \\ &+ \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ C\bar{G} & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots \\ C\bar{A}^{N_d-2}\bar{G} & \cdots & C\bar{G} & 0 \end{bmatrix}}_{\Gamma_f} \begin{bmatrix} \bar{Q}_w & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{Q}_w \end{bmatrix} \Gamma_f^T + \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} \Psi^T\end{aligned} \tag{4.24}$$

where,

$$\Psi = \Gamma_f \begin{bmatrix} -AL & 0 & 0 & 0 \\ 0 & -AL & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & -AL \end{bmatrix}, \qquad \bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}$$

Following Equation 4.23, we see that ε_k is a linear combination of normally distributed noises given $\bar{A}^k \approx 0$ and hence is normal. This implies $\mathcal{Y}_k$ is also normally distributed. We then have:

$$Y_p \sim N(0, \Omega_p)$$

Next we use Equation 4.22 and the above result to get the distribution of $\mathbb{Y}$. Since $\mathbb{Y}$ is a matrix, its mean and covariance are defined for the stacked version of the matrix

i.e. $\mathbb{Y}_s$. Given the linear relationship between Y_p and $\mathbb{Y}_s$, we get,

$$\mathbb{Y}_s \sim N(0, \mathbb{E}_1 \Omega_p \mathbb{E}_1^T)$$

Thus, the covariance of $\mathbb{Y}$ is:

$$\Omega = \mathbb{E}_1 \Omega_p \mathbb{E}_1^T \tag{4.25}$$

where Ω_p is given by Equation 4.24 and $\mathbb{E}_1$ is defined in Equation 4.22.

□

Chapter 5

Applying the ALS Technique for Misassigned Disturbance Models in Offset-free Model Predictive Control [1]

Model Predict ive Control (MPC) has emerged as one of the most important advanced control methods in the chemical industry (Qin and Badgwell, 2003). MPC casts the control problem in the form of an optimization, which makes it convenient to handle constraints and nonlinear models explicitly. Most current formulations of MPC have a state estimator along with the dynamic constrained regulator and a target calculator. Nominal asymptotic stability for a deterministic linear MPC was shown by Rawlings and Muske (1993) under the assumption that the model is known perfectly and there is full state information. With random disturbances however, the state is no longer

[1] **Portions of this chapter are to appear in Rajamani et al. (August, 2007)**

known exactly and one requires dynamic feedback from the measurements to update the current estimate of the state. The function of the state estimator is to reconcile the past and current measurements to estimate the current state of the system. Nominal stability of the state estimator combined with the regulator was shown by Scokaert et al. (1997). The regulator and the target calculator require only estimates of the current state and the disturbance to make a prediction of optimal control moves over a future horizon.

Any correlated disturbance, that enters the plant and is not included in the model, results in biased predicted states. It is customary to assume that the disturbance enters through the inputs or the outputs. The bias can be removed by using the predicted and measured outputs to estimate the disturbance. In the target calculator formulation, the state is augmented with integrating disturbances. The target calculator then ensures offset free control in presence of unmodelled disturbances (Muske and Rawlings, 1993a) by shifting the steady state target for the regulator depending on the estimate of the bias (with the assumption that the bias remains constant in the future). The integrating disturbance can be added either to the input (Davison and Smith, 1971; Morari and Stephanopoulos, 1980), to the output (Richalet et al., 1978; Cutler and Ramaker, 1980) or a combination of both (Marquis and Broustail, 1988; Pannocchia and Rawlings, 2002; Muske and Badgwell, 2002). Most industrial MPC implementations add the bias term to the output (Qin and Badgwell, 2003; Downs, 1999). Pannocchia and Rawlings (2002) and Muske and Badgwell (2002) independently derived rank conditions that should be

satisfied for the disturbance models to ensure offset-free control. They also point out the lack of consensus in the literature on the choice of the disturbance model. Their examples show significant difference in the closed-loop behavior of the controller depending on the choice of the disturbance model. Recently Pannocchia and Bemporad (2007) present a method for a combined design of a disturbance model and its observer by solving an appropriate H_∞ control problem. Their work, however, neglects stochastic disturbances that enter the state and the measured outputs.

The main contribution of this chapter is to show that for linear models, the choice of the disturbance model does not affect the performance of the closed-loop when the estimator gain is found from appropriately chosen noise covariances. The equivalence in the estimator gains is shown by appealing to realization theory (Ho and Kalman, 1966). Subsequently the performance of any regulator that is a feedback on the current estimate of the state remains unaffected by the choice of the disturbance model. The covariances can be estimated from data using autocovariances at different time lags using the ALS technique developed in Chapter 3. The state estimator gain thus determined from the estimated covariances gives optimal closed-loop performance.

The rest of the chapter is organized as follows: In Section 5.1 the equations for estimation, regulation and target calculation in MPC are given. We motivate the rest of the chapter by examining the consequences of an incorrect disturbance model in Section 5.2. Section 5.3 formulates the equivalence transformation between arbitrary choices of the disturbance models. Section 5.4 presents the autocovariance least-squares (ALS)

technique for estimating the covariances for incorrectly chosen disturbance models followed by examples in section 5.5. Finally, conclusions are presented in Section 5.6.

5.1 Background

We start with the following discrete-time state-space description of the model as given in Chapter 1:

Plant:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Gw_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{5.1}$$

with the usual notations and dimensions. Let the covariances of w_k, v_k be Q_w, R_v respectively and also assume that w_k, v_k are not cross correlated as in Chapter 3. The pair (A, C) is assumed to be observable and the pair (A, B) assumed controllable. Let y_k^c be the set of controlled outputs that are a subset of the measurements y_k given by $y_k^c = Hy_k$. In the rest of this chapter we assume without loss of generality that the set of controlled outputs are the same as the measured outputs with $H = I$. We also use the term "measurements" and "outputs" interchangeably to mean y_k. The analysis in this chapter remains the same when y_k^c is only a subset of y_k.

As before, we assume that the deterministic parameters (A, B, C) of the plant in Equation 5.1 are known or estimated using identification procedures and the stochastic parameters G, Q_w, R_v are unknown. Even though the deterministic parameters may be known perfectly, unmodelled disturbances can enter through the inputs or the outputs

and result in an offset in the controller. A physical example of such a disturbance is a leak in the feed pipe causing decreased input flow rate or cooling liquid contamination decreasing the effective cooling rate. Such a disturbance can often be adequately described as a integrating white noise with some small unknown covariance Q_ξ for ξ_k:

Unmodelled Disturbance in the plant:

$$d_{k+1} = d_k + \xi_k \tag{5.2}$$

Figure 5.1 shows an example of an integrating disturbance following Equation 5.2 with covariance $Q_\xi = 0.0008$.

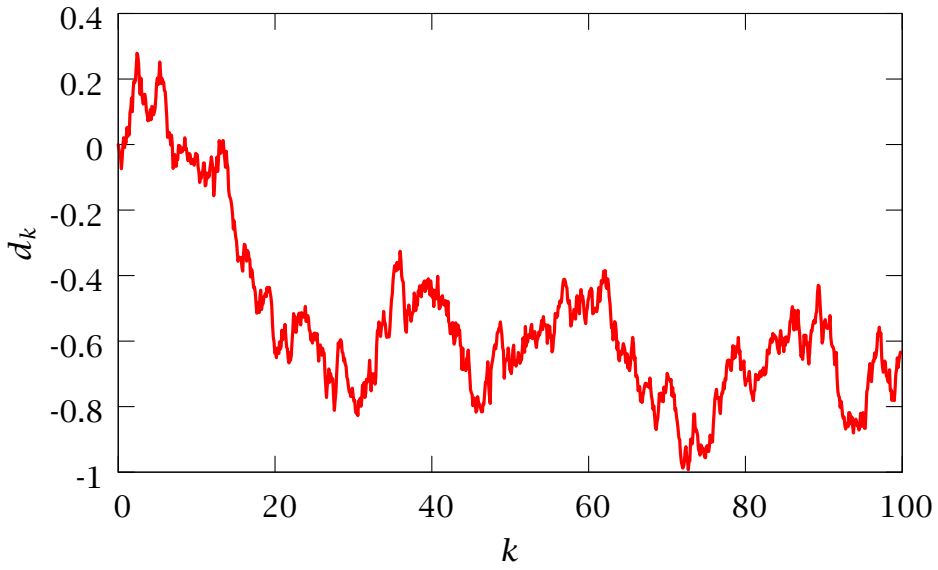


Figure 5.1: Example of slow unmodelled disturbance entering the plant in the input or the output

To ensure offset-free control in the presence of unmodelled disturbances Model Predictive Control (MPC) models add an integrating disturbance modelled at either the

inputs or the outputs of the system (Francis and Wonham, 1975, 1976). Let $d_k \in \mathbb{R}^{n_d}$ be the integrating disturbance entering the state through the matrix $B_d \in \mathbb{R}^{n \times n_d}$ and the measurement through $C_d \in \mathbb{R}^{p \times n_d}$. The state x in the model is augmented with the integrating disturbance.

Augmented Model:

$$\begin{aligned} \begin{bmatrix} x \\ d \end{bmatrix}_{k+1} &= \underbrace{\begin{bmatrix} A & B_d \\ 0 & I \end{bmatrix}}_{A_a} \begin{bmatrix} x \\ d \end{bmatrix}_k + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k + \underbrace{\begin{bmatrix} G & 0 \\ 0 & I \end{bmatrix}}_{G_a} \begin{bmatrix} w \\ \xi \end{bmatrix}_k \\ y_k &= \underbrace{\begin{bmatrix} C & C_d \end{bmatrix}}_{C_a} \begin{bmatrix} x \\ d \end{bmatrix}_k + v_k \end{aligned} \tag{5.3}$$

The choice of the structure of the disturbance B_d, C_d is usually unknown and a general unstructured disturbance model can have disturbances entering both the states and the outputs. The conditions on the models B_d, C_d for detectability and to ensure offset free control were proved by Pannocchia and Rawlings (2002) and Muske and Badgwell (2002).

In addition to the covariances GQ_wG^T, R_v, the covariance Q_ξ of the disturbance ξ_k is also required now to specify the state estimator gain. Usual choices for the disturbance model are (Muske and Rawlings, 1993a):

1. $B_d = 0, C_d = I$ called the output disturbance model, or
2. $B_d = B, C_d = 0$ called the input disturbance model

The number of disturbances n_d has to be chosen as $n_d = p$ to ensure zero offset in all

the outputs (Pannocchia and Rawlings, 2002).

With Equation 5.1 representing the plant, Equation 5.2 for the unmodelled disturbance in the plant and the augmented model in Equation 5.3, the state estimation and regulation in MPC are described in the following subsections. The basic block diagram

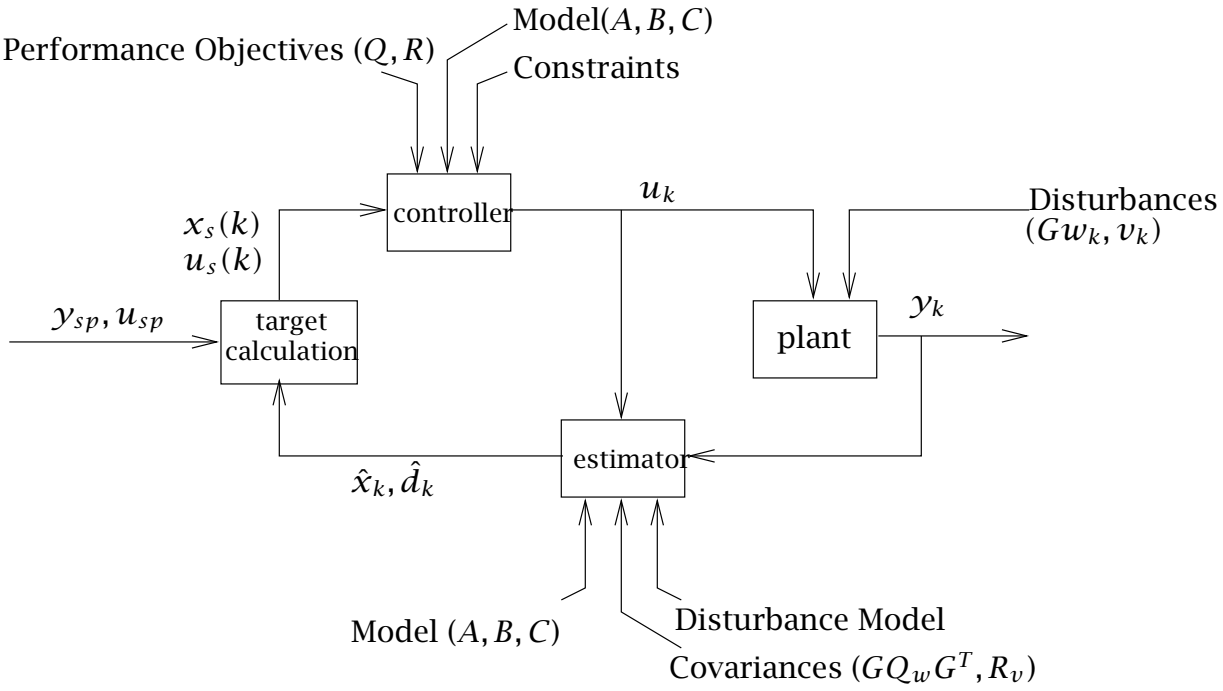


Figure 5.2: Block diagram representing parts of the Model Predictive Control (MPC) scheme

for the linear MPC scheme is shown in Figure 5.2.

5.1.1 State Estimation

The state estimation part of MPC provides feedback from the outputs. Given a set of measurements $\{y_0, y_1, \cdots, y_k\}$ and inputs $\{u_0, u_1, \cdots, u_k\}$ up to time t_k, the optimal

state estimation for the model in Equation 5.3 with no constraints is given by the Kalman filter. For linear time-invariant models like the ones considered in this chapter, the Kalman filter is the set of simple recursive calculations shown below. Let $\hat{x}_{k|k}$ be the filtered estimate of the state x_k given data up to t_k and $\hat{x}_{k|k-1}$ be the predicted state given data up to t_{k-1}. The prediction equations are:

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k} + Bu_k + B_d\hat{d}_{k|k} \\ \hat{d}_{k+1|k} &= \hat{d}_{k|k} \end{aligned} \tag{5.4}$$

and the filtering equations are:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + L_x(y_k - C\hat{x}_{k|k-1} - C_d\hat{d}_{k|k-1}) \\ \hat{d}_{k|k} &= \hat{d}_{k|k-1} + L_d(y_k - C\hat{x}_{k|k-1} - C_d\hat{d}_{k|k-1}) \end{aligned} \tag{5.5}$$

where L_x and L_d are the steady state Kalman filter gains calculated using covariances GQ_wG^T, Q_ξ and R_v and using the Riccati equation for the augmented system defined in Equation 5.3. For a time-invariant model the steady state Kalman gains L_x, L_d are also time-invariant.

$$\begin{aligned} P_a &= A_aP_aA_a^T - A_aP_aC_a^T(C_aP_aC_a^T + R_v)^{-1}C_aP_aA_a^T + \begin{bmatrix} GQ_wG^T & 0 \\ 0 & Q_\xi \end{bmatrix} \\ L_a &= P_aC_a^T(C_aP_aC_a^T + R_v)^{-1} \end{aligned} \tag{5.6}$$

in which, $L_a = [L_x^T \quad L_d^T]^T$.

Other advanced state estimators, e.g. the Moving Horizon Estimator (MHE) are required when the model is nonlinear or the system has constraints (Rao et al., 2001).

5.1.2 Target Calculation

Note that the disturbance d_k in the augmented model is uncontrollable. The use of d_k is however allows removal of offset in all the outputs y_k when there are unmodelled disturbances. This follows from the "internal model principle" for linear multivariable systems as discussed in Francis and Wonham (1976). The target calculator is then used to calculate a new steady state target x_s, u_s by solving an optimization based on the estimate $\hat{d}_{k|k}$.

If y_{sp} and u_{sp} are the set points for the output and the input and R_s is a positive definite quadratic penalty, the modified steady state targets $x_s \in \mathbb{R}^n, u_s \in \mathbb{R}^m$ are calculated at each sampling time as a solution to the following convex quadratic programming problem as shown in Muske and Rawlings (1993a):

$$\min_{x_s, u_s} \quad (u_s - u_{sp})^T R_s (u_s - u_{sp}) \tag{5.7a}$$

subject to,

$$\begin{bmatrix} I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix} = \begin{bmatrix} B_d \hat{d}_{k|k} \\ -C_d \hat{d}_{k|k} + y_{sp} \end{bmatrix} \tag{5.7b}$$

and inequality constraints on x_s, u_s

The above optimization has a feasible solution whenever the inequality constraints in 5.7b are not active. The objective function 5.7a is modified when there is no feasible solution (Muske and Rawlings, 1993a).

5.1.3 Regulation

The receding horizon formulation for the regulator in MPC is given as the quadratic objective function with the model constraints and penalties $Q \geq 0, R \geq 0$. The state and the input are penalized for deviations from the steady state targets calculated from Equation 5.7a (Muske and Rawlings, 1993a).

$$\begin{aligned} \min_{u_k,\cdots,u_{k+N_c}} \quad & \sum_{j=k}^{k+N_c} (\hat{x}_{j|k} - x_s)^T Q (\hat{x}_{j|k} - x_s) + (u_j - u_s)^T R (u_j - u_s) \\ \text{subject to,} \quad & \hat{x}_{j+1|k} = A\hat{x}_{j|k} + Bu_j + B_d\hat{d}_{k|k}, \quad j \geq k \end{aligned} \tag{5.8}$$

The state and input constraints are not shown in the above formulation for ease of notation. Inclusion of the constraints do not change any of the analysis that follows. Here, the control horizon and the prediction horizon are taken to be the same: N_c. Assuming that the filtered disturbance $\hat{d}_{k|k}$ remains unchanged over the future horizon N_c and the current state taken as the estimate $\hat{x}_{k|k}$, the optimized control moves $\{u_k^*, u_{k+1}^*, \cdots u_{k+N_c}^*\}$ over the horizon are calculated and only the first move u_k^* is implemented at time t_k.

The estimation, target calculation and regulation is repeated at every time step. For $N_c \to \infty$ and no constraints on input and state, the optimization in Equation 5.8 is equivalent to the Linear Quadratic Regulator (LQR).

Remark 5.1. *Note that the state estimation in Section 5.1.1 requires knowledge of the covariances GQ_wG^T, Q_ξ, R_v and the target calculation in Section 5.1.2 requires the structure of the unmodelled disturbances given by the matrices B_d, C_d. In general, none of*

these are known from first principles modelling. The ideal choice of B_d, C_d is one that represents the unmodelled disturbance in the plant accurately. The choice of the disturbance model has been the subject of research (Lee et al., 1994; Pannocchia, 2003; Eppler, 1997; Shinskey, 1994, chap. 5). These studies assume that the covariances GQ_wG^T, Q_ξ, R_v are known and the models B_d, C_d are unknown. However in Sections 5.3 and 5.4 we show that if an incorrect choice for B_d, C_d is made, then the covariances GQ_wG^T, Q_ξ, R_v estimated from data are within an equivalent transformation such that the particular choice of B_d, C_d has no effect on the closed-loop performance.

5.2 Motivation

The rules for choosing the penalties Q, R for the regulator are well researched and based on intuitive reasoning (see for example Anderson and Moore (1990, chap. 6)). The estimator on the other hand requires knowledge of the disturbance covariances GQ_wG^T, R_v, Q_ξ and the disturbance models B_d, C_d. Consider an unconstrained single input single output plant with the following values for the parameters:

$$A = 0.5, \quad B = 0.5, \quad C = 1, \quad y_{sp} = 0.8$$

Let a deterministic unmodelled disturbance of magnitude 1 enter the input at sampling time 3 and is measured at sampling time 4. Let there be no other noises affecting the plant: $GQ_wG^T = 0, R_v = 0$. An incorrect choice of the MPC disturbance model would be $B_d = 0, C_d = 1$ modelling the disturbance as entering at the output. The actual plant

model on the other hand is $B_d = B = 0.5, C_d = 0$ (the input disturbance model).

We now compare the closed-loop performance for the incorrect disturbance model against the plant disturbance model. The regulator is chosen to be the LQR with penalties $Q = 1, R = 0$ and penalty $R_s = 1$ for the target calculator. If the covariances GQ_wG^T, Q_ξ, R_v were assumed to be known then the optimal state estimator is the deadbeat Kalman filter gain for the input disturbance model calculated by substituting $GQ_wG^T = 0, R_v \approx 0, Q_\xi \neq 0$ in the Kalman filter Equation 5.6. The deadbeat Kalman gains for the input disturbance model is $L_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and for the incorrect output disturbance model is $L_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. The control law is the same both cases and given by the LQR feedback: $u_k = K(\hat{x}_{k|k} - x_s)$. The rejection of the deterministic disturbance in the two cases is show in Figure 5.3. With the input disturbance model, the disturbance in the input is estimated and rejected much more quickly than when using the incorrect model although the same covariances GQ_wG^T, R_v, Q_ξ are used in the calculation of both the state estimators. The sluggish behavior of the output disturbance model and the decrease in closed-loop performance is due to the choice of the disturbance model and the chosen estimator gain. A similar difference in the performance with the incorrect choice of the disturbance model was noted by Shinskey (1994, chap. 5); Lee et al. (1994); Muske and Rawlings (1993b); Muske and Badgwell (2002).

Since we usually lack knowledge about the structure of the unmodelled disturbances (matrices B_d, C_d), the only other handle to change the input-output performance

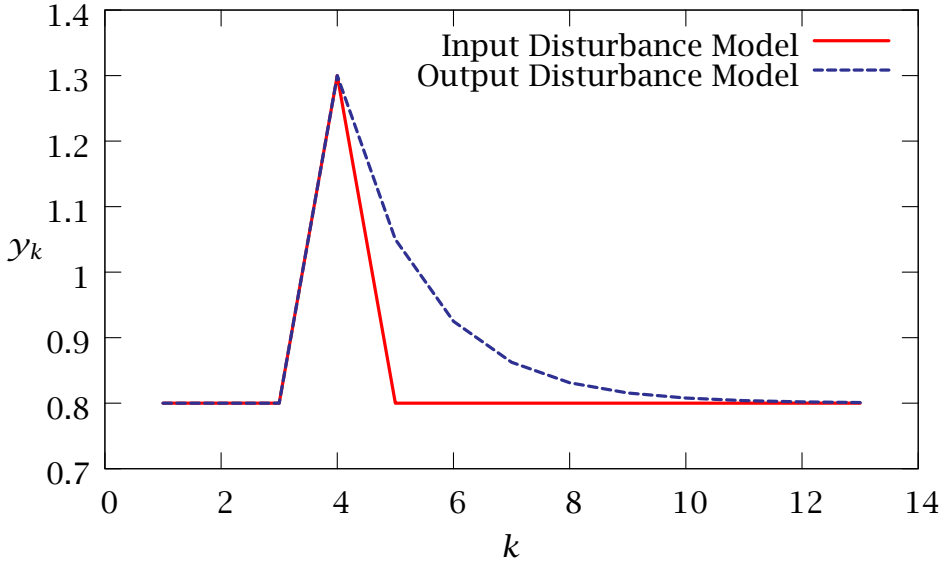


Figure 5.3: Disturbance rejection using two disturbance models

is the estimator gain L. Once the choice of the disturbance model is fixed, we can use a data-based technique that estimates $\hat{L}$ such that the closed-loop performance is not affected by the choice of the disturbance model (Section 5.4). For example, if in the above problem, the choice of the disturbance model is already fixed incorrectly as the output disturbance model, then a transformation of the covariances $\widehat{GQ_wG^T}, \hat{Q}_\xi, \hat{R}_\nu$ can be estimated from data to give the same closed-loop performance as achieved using the input disturbance model and the original covariances.

5.3 Model Equivalence

The choice of the disturbance model for offset-free control in MPC is often ambiguous with no real guidelines available. In Pannocchia (2002) a min-max problem is solved to

find the best performing disturbance model. Most industrial MPC implementations such as IDCOM (Richalet et al., 1978), DMC (Cutler and Ramaker, 1980) and QDMC (García and Morshedi, 1986) use an output disturbance model to remove offset . The choice of the output disturbance model has been criticized in the literature as being sluggish in rejecting disturbances (Shinskey, 1994, chap. 5),(Lee et al., 1994). In this section we show that for linear time-invariant (LTI) systems, disturbance models are equivalent as long as the estimator is determined with the appropriate covariances.

Remark 5.2. *It is interesting to examine some of the early history of disturbance modelling. Box and Jenkins clearly make the point that one can identify equivalent disturbance models that represent different disturbance sources by what they call "transfer of disturbance origins." (Box and Jenkins, 1970, p. 469, 1st edition). Note that the interested reader must obtain the first edition because by the 3rd edition (Box et al., 1994), this terminology and entire discussion had been deleted. Box and Jenkins also point out that identifying a disturbance model with data collected in a fashion that does not include all relevant disturbances that are encountered in closed-loop operation does not produce an optimal control because the disturbance model consequently has significant error. They recommend using closed-loop identification (Box and Jenkins, 1970, p. 471, 1st ed) to avoid the accidental neglect of disturbance sources.*

In this work, we are emphasizing again Box and Jenkins's conclusion that one is free to choose the source of the disturbance in the disturbance model, and one naturally requires data that include all relevant disturbances. It is irrelevant, however, if these data

are collected under closed-loop or open-loop control. Also, we are not interested in identifying the disturbance model, which was the primary focus of Box and Jenkins's earlier work. We fully expect the practitioner to employ an integrated white noise disturbance model to represent essentially all of the plant disturbances. The motivation is to remove steady offset in the closed-loop system output, not to capture an accurate dynamic model of the true disturbance. A contribution of the current chapter therefore is to propose and show that such rough disturbance models are often completely adequate to handle disturbances entering from other sources with different dynamics, **if one estimates the statistics of the driving noise from data**.

Consider two choices of the disturbance models (B_{d1}, C_{d1}) and (B_{d2}, C_{d2}) satisfying detectability conditions. If n_d is the dimension of the disturbance vector, the assumption of detectability implies (Pannocchia and Rawlings, 2002):

$$\text{rank}\begin{bmatrix} A-I & B_{d1} \\ C & C_{d1} \end{bmatrix} = n + n_d, \quad \text{rank}\begin{bmatrix} A-I & B_{d2} \\ C & C_{d2} \end{bmatrix} = n + n_d \tag{5.9}$$

Note that if the disturbance in the actual plant is not detectable, then the state estimator is defined for a transformed, lower dimensional model that is detectable.

The systems augmented with the disturbance model are simple nonminimal realizations of the same unaugmented model where d_k is an uncontrollable state. The augmented systems with the two general disturbance models can be written as follows:

Model 1:

$$\begin{aligned}\begin{bmatrix} x^1 \\ d^1 \end{bmatrix}_{k+1} &= \underbrace{\begin{bmatrix} A & B_{d1} \\ 0 & I \end{bmatrix}}_{A_1} \begin{bmatrix} x^1 \\ d^1 \end{bmatrix}_k + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{B_a} u_k + G_1 \begin{bmatrix} w \\ \xi \end{bmatrix}_k \\ y_k &= \underbrace{\begin{bmatrix} C & C_{d1} \end{bmatrix}}_{C_1} \begin{bmatrix} x^1 \\ d^1 \end{bmatrix}_k + v_k \end{aligned} \tag{5.10}$$

Model 2:

$$\begin{aligned}\begin{bmatrix} x^2 \\ d^2 \end{bmatrix}_{k+1} &= \underbrace{\begin{bmatrix} A & B_{d2} \\ 0 & I \end{bmatrix}}_{A_2} \begin{bmatrix} x^2 \\ d^2 \end{bmatrix}_k + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{B_a} u_k + G_2 \begin{bmatrix} w \\ \xi \end{bmatrix}_k \\ y_k &= \underbrace{\begin{bmatrix} C & C_{d2} \end{bmatrix}}_{C_2} \begin{bmatrix} x^2 \\ d^2 \end{bmatrix}_k + v_k \end{aligned} \tag{5.11}$$

If the two models have the same input to output relationship, there exists a similarity transformation matrix T such that:

$$TA_1T^{-1} = A_2, \qquad C_1T^{-1} = C_2, \qquad TB_a = B_a \tag{5.12}$$

Also for Models 1 and 2 the unknown stochastic parameters G_1 and G_2 must satisfy:

$$TG_1 = G_2 \tag{5.13}$$

Lemma 5.1. *If the augmented models in Equations 5.10 and 5.11 are assumed to be detectable, then the transformation relations in Equation 5.12 can be expressed equivalently*

as:

$$T_{11} = I, \qquad \begin{bmatrix} A - I & B_{d2} \\ C & C_{d2} \end{bmatrix} \begin{bmatrix} T_{12} \\ T_{22} \end{bmatrix} = \begin{bmatrix} B_{d1} \\ C_{d1} \end{bmatrix}, \qquad T_{21} = 0$$

where, T is partitioned as (with appropriate dimensions):

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$$

Proof. The proof is given in Appendix 5.7.1 □

Theorem 5.1. *For the two detectable augmented models in Equations 5.10 and 5.11, a transformation matrix T satisfying the relations in Equation 5.12 exists and is unique if and only if the following condition holds:*

$$rank \begin{bmatrix} A - I & B_{d1} & B_{d2} \\ C & C_{d1} & C_{d2} \end{bmatrix} = n + n_d \tag{5.14}$$

Proof. The proof follows from the relations in Equation 5.12 expressed equivalently from Lemma 5.1 and the detectability conditions in Equation 5.9. □

Note that for zero offset in all the outputs y_k, we need $n_d = p$. The condition in Theorem 5.1 is then automatically satisfied for all detectable disturbance models.

Remark 5.3. *Notice that we are considering a known and fixed minimal input-output realization of the triple (A, B, C). The LQR cost in Equation 5.8 then remains unchanged for the two nonminimal augmented models in Equations 5.10 and 5.11 related by the conditions given in Lemma 5.1.*

An alternate derivation and condition for the transformation matrix in Equation 5.12 is given in Appendix 5.7.3.

5.3.1 Special Case of Equivalence

For an input disturbance model we have $B_d = B$ and $C_d = 0$ and for an output disturbance model we have $B_d = 0$ and $C_d = I$. For these particular choices of the disturbance models with $n_d = m = p$, we get the following transformation matrix:

- Model 1- $B_{d1} = B, C_{d1} = 0$, Model 2-$B_{d2} = 0, C_{d2} = I$

$$T = \begin{bmatrix} I & (A-I)^{-1}B \\ 0 & -C(A-I)^{-1}B \end{bmatrix} \tag{5.15}$$

Remark 5.4. *Note that when A has an integrator, the inverses above are not defined. In this case the output disturbance model is not detectable and fails Theorem 5.1. This has been long known in the literature as a potential problem with industrial MPC implementations such as IDCOM, DMC, QDMC, which use a constant output step disturbance model (Morari and Lee, 1991; Muske and Badgwell, 2002; Qin and Badgwell, 2003). Various fixes have also been suggested in the literature by Marquis and Broustail (1988); Li et al. (1989); Lundström et al. (1995); Gupta (1998) and Dougherty and Cooper (2003).*

5.3.2 Equivalent Closed-loop Performance

Given Models 1 and 2 related by the equivalence transformation matrix T satisfying the relations in Equation 5.12, from realization theory (Ho and Kalman, 1966) equiv-

alent closed-loop performance for the two models implies that given a sequence of control actions $\{u_1, \cdots, u_{k-1}\}$ and noise sequences $\{w_1, \cdots, w_{k-1}\}$, $\{\xi_1, \cdots \xi_{k-1}\}$ and $\{v_1, \cdots, v_k\}$, then the output sequence from the two models: $\{y_1, \cdots, y_k\}$ are the same starting from zero initial conditions. When these models are used in MPC, the noises w_k, ξ_k, v_k are unknown.

If L_1 and L_2 are the state estimator gains (not necessarily the Kalman filter gains) for Models 1 and 2, then equivalent closed-loop performance implies that given the inputs $\{u_1, \cdots, u_{k-1}\}$ and outputs $\{y_1, \cdots, y_k\}$, the innovations sequence $\{\mathcal{Y}_1, \cdots, \mathcal{Y}_k\}$ are the same, where $\mathcal{Y}_k = (y_k - \hat{y}_{k|k-1})$. The innovations for the augmented system in Equation 5.3 are given by Equations 5.4 and 5.5:

$$\begin{aligned} \begin{bmatrix} \hat{x} \\ \hat{d} \end{bmatrix}_{k+1|k} &= A_a \begin{bmatrix} \hat{x} \\ \hat{d} \end{bmatrix}_{k|k-1} + B_a u_k + A_a L_a (y_k - \hat{y}_{k|k-1}) \\ \mathcal{Y}_k &= y_k - C_a \begin{bmatrix} \hat{x} \\ \hat{d} \end{bmatrix}_{k|k-1} \end{aligned} \tag{5.16}$$

Lemma 5.2. *Given a set of measurements $\{y_1, \cdots, y_k\}$ and inputs $\{u_1, \cdots, u_{k-1}\}$ and the two disturbance models in Equations 5.10 and 5.11 related to each other through the unique transformation matrix T satisfying Equation 5.12, the innovations sequence defined as $\mathcal{Y}_k = (y_k - \hat{y}_{k|k-1})$ are the same for the two models if and only if the state estimator gains (can be suboptimal gains) satisfy the relation:*

$$L_2 = TL_1 \tag{5.17}$$

Proof. The proof follows using simple algebraic manipulations using Equations 5.10

and 5.11 in Equation 5.16 and the transformations in Equation 5.12. □

Remark 5.5. *Since Lemma 5.2 makes no assumptions on the optimality of the estimator gains L_1 and L_2, the result also holds when the optimal Kalman filter gains are used instead.*

For the Kalman filter, the calculation of the estimator gains follow from the Riccati Equation 5.6.

For Model 1:

$$\begin{aligned} P_1 &= A_1 P_1 A_1^T - A_1 P_1 C_1^T (C_1 P_1 C_1^T + R_v)^{-1} C_1 P_1 A_1^T + G_1 Q_a G_1^T \\ L_1 &= P_1 C_1^T (C_1 P_1 C_1^T + R_v)^{-1} \end{aligned} \tag{5.18}$$

and for Model 2:

$$\begin{aligned} P_2 &= A_2 P_2 A_2^T - A_2 P_2 C_2^T (C_2 P_2 C_2^T + R_v)^{-1} C_2 P_2 A_2^T + G_2 Q_a G_2^T \\ L_2 &= P_2 C_2^T (C_2 P_2 C_2^T + R_v)^{-1} \end{aligned} \tag{5.19}$$

in which,

$$Q_a = \begin{bmatrix} Q_w & 0 \\ 0 & Q_\xi \end{bmatrix}$$

To obtain optimal closed-loop performance, we need to know the transformation matrix T relating the plant disturbance model to the controller disturbance model in addition to the original covariances Q_a, R_v. Since the disturbance in the plant is unmodelled and the source unknown, the transformation matrix T is also unknown. Hence, the transformation in the state estimator gains as shown in Equation 5.12 cannot be

used. However, using Equations 5.18 and 5.19 we see that if covariances $G_1 Q_a G_1^T, R_v$ for Model 1 and $G_2 Q_a G_2^T, R_v$ for Model 2 are used to calculate the Kalman filter, where $TG_1 = G_2$, then Equation 5.12 is satisfied.

Thus, if the covariances $G_1 Q_a G_1^T, R_v$ can be estimated from data, then without knowing the transformation matrix T, Model 1 with estimator gain L_1 performs equivalently in closed-loop as Model 2 with estimator gain L_2. The choice of the disturbance model is then irrelevant to the closed-loop input-output behavior.

5.4 Using Correlations to Estimate Filter Gain

Again consider the plant given by a Linear Time-Invariant (LTI) state-space model as shown in Equation 5.1:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Gw_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{5.20}$$

Let d_k be an unmodelled integrating white noise disturbance entering the plant through the disturbance model structures B_{dp}, C_{dp}. The augmented plant model is then given by:

$$\begin{aligned}
\begin{bmatrix} x^p \\ d^p \end{bmatrix}_{k+1} &= \underbrace{\begin{bmatrix} A & B_{dp} \\ 0 & I \end{bmatrix}}_{A_p} \begin{bmatrix} x^p \\ d^p \end{bmatrix}_k + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{B_p} u_k + \underbrace{\begin{bmatrix} G & 0 \\ 0 & I \end{bmatrix}}_{G_p} \begin{bmatrix} w \\ \xi \end{bmatrix}_k \\
y_k &= \underbrace{\begin{bmatrix} C & C_{dp} \end{bmatrix}}_{C_p} \begin{bmatrix} x^p \\ d^p \end{bmatrix}_k + v_k
\end{aligned} \tag{5.21}$$

Since knowledge of the disturbance structure is unknown, assume the disturbance model in MPC to remove offset is incorrectly chosen to have the structure B_{dm}, C_{dm}:

$$\begin{aligned}
\begin{bmatrix} x^m \\ d^m \end{bmatrix}_{k+1} &= \underbrace{\begin{bmatrix} A & B_{dm} \\ 0 & I \end{bmatrix}}_{A_m} \begin{bmatrix} x^m \\ d^m \end{bmatrix}_k + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{B_p} u_k + G_m \begin{bmatrix} w \\ \xi \end{bmatrix}_k \\
y_k &= \underbrace{\begin{bmatrix} C & C_{dm} \end{bmatrix}}_{C_m} \begin{bmatrix} x^m \\ d^m \end{bmatrix}_k + v_k
\end{aligned} \tag{5.22}$$

The superscripts p and m are used to denote states for the actual plant and the model respectively. Assume that the conditions in Theorem 5.1 are satisfied and there exists an *unknown* transformation matrix T that connects models in Equations 5.21 and 5.22 through the relations in Equation 5.12. Then the following holds from realization theory:

$$T \begin{bmatrix} x^p \\ d^p \end{bmatrix}_k = \begin{bmatrix} x^m \\ d^m \end{bmatrix}_k, \quad TA_pT^{-1} = A_m, \quad C_pT^{-1} = C_m \tag{5.23}$$

We assume (A, B, C) to be known from some system identification scheme, the stochastic parameters G, Q_w, Q_ξ, R_v are unknowns and B_{dm}, C_{dm} are incorrectly chosen.

Given some arbitrary (stable, perhaps suboptimal) initial estimator L, the state estimates given the incorrect model are given by:

$$\begin{aligned}\begin{bmatrix}\hat{x}^m\\ \hat{d}^m\end{bmatrix}_{k+1|k} &= A_m\begin{bmatrix}\hat{x}^m\\ \hat{d}^m\end{bmatrix}_{k|k-1} + B_p u_k + A_m L\left(y_k - C_m\begin{bmatrix}\hat{x}^m\\ \hat{d}^m\end{bmatrix}_{k|k-1}\right)\\ \hat{y}_{k|k-1} &= C_m\begin{bmatrix}\hat{x}^m\\ \hat{d}^m\end{bmatrix}_{k|k-1}\end{aligned} \tag{5.24}$$

If the innovations sequence are given by $\gamma_k = y_k - \hat{y}_{k|k-1}$, using Equations 5.21 and 5.24, we get:

$$\gamma_k = C_p\begin{bmatrix}x^p\\ d^p\end{bmatrix}_k - C_m\begin{bmatrix}\hat{x}^m\\ \hat{d}^m\end{bmatrix}_{k|k-1} + v_k$$

If the estimate error is defined as $\varepsilon_k = \begin{bmatrix}x_k^m - \hat{x}_{k|k-1}^m\\ d_k^m - \hat{d}_{k|k-1}^m\end{bmatrix}$, substituting from Equation 5.23, we get:

$$\gamma_k = C_m\varepsilon_k + v_k$$

We can then write the evolution of the state estimate error by subtracting the plant Equation 5.21 from the model in Equation 5.24 and using the transformations in

Equation 5.23:

$$\varepsilon_{k+1} = \underbrace{(A_m - A_m L C_m)}_{\bar{A}} \varepsilon_k + \underbrace{\begin{bmatrix} TG_p & -A_m L \end{bmatrix}}_{\bar{G}} \begin{bmatrix} w \\ \xi \\ v \end{bmatrix}_k \tag{5.25}$$

$$y_k = C_m \varepsilon_k + v_k$$

We define the autocovariance as $C_j = E[y_k y_{k+j}^T]$, the expectation of data at different time lags Jenkins and Watts (1968, p. 146). Using Equation 5.25 we can write the following expectation of covariances:

$$E[y_k y_k^T] = C_m P C_m^T + R_v \tag{5.26}$$

$$E[y_{k+j} y_k^T] = C_m \bar{A}^j P C_m^T - C_m \bar{A}^{j-1} A_m L R_v \quad j \geq 1 \tag{5.27}$$

which are independent of k when we assume that data are collected when operating at steady state. Again using Equation 5.25 we note that $P = E[\varepsilon_k \varepsilon_k^T]$ the steady state estimate error covariance satisfies the Lyapunov equation:

$$P = \bar{A} P \bar{A}^T + \underbrace{\begin{bmatrix} TG_p & -A_m L \end{bmatrix}}_{\bar{G}} \begin{bmatrix} Q_a & 0 \\ 0 & R_v \end{bmatrix} \bar{G}^T \qquad \text{where, } Q_a = \begin{bmatrix} Q_w & 0 \\ 0 & Q_\xi \end{bmatrix} \tag{5.28}$$

A computation of the first column block of the Autocovariance Matrix of innovations process over some user defined window N then gives (Chapter 3 and Odelson

et al. (2006b)):

$$\mathcal{R}_1(N) = \begin{bmatrix} C_0 \\ C_1 \\ \vdots \\ C_{N-1} \end{bmatrix} = \underbrace{\begin{bmatrix} C_m \\ C_m\bar{A} \\ \vdots \\ C_m\bar{A}^{N-1} \end{bmatrix}}_{\mathcal{O}} PC_m^T + \underbrace{\begin{bmatrix} I \\ -C_m A_m L \\ \vdots \\ -C_m\bar{A}^{N-2}A_m L \end{bmatrix}}_{\Gamma} R_v \tag{5.29}$$

Define the following covariance product:

$$\mathbb{Q} \triangleq TG_pQ_aG_p^TT^T$$

Notice that in the above Equation 5.29 with P given by Equation 5.28, the only unknowns are the transformed covariance $\mathbb{Q}$ and R_v . If $\hat{\mathcal{R}}_1(N)$ is an estimate of $\mathcal{R}_1(N)$ from data, then similar to the development of the ALS technique in Chapter 3, we can express Equation 5.29 as a weighted least-squares problem in a vector of unknowns $(\mathbb{Q})_s, (R_v)_s$. Here, the subscript 's' denotes the column-wise stacking of the elements of the matrix in a vector. As before, we add semidefinite constraints to ensure the covariances to be positive semidefinite. We then get the following convex autocovariance least-squares (ALS) optimization problem developed in Chapter 3:

$$\Phi = \min_{\mathbb{Q},R_v} \left\| \mathcal{A} \begin{bmatrix} (\mathbb{Q})_s \\ (R_v)_s \end{bmatrix} - b \right\|_W^2 \tag{5.30}$$

$$\text{subject to, } \mathbb{Q} \geq 0, \quad R_v \geq 0, \quad \mathbb{Q} = \mathbb{Q}^T, \quad R_v = R_v^T$$

where $\mathcal{A}$ is defined from Chapter 3 as,

$$\begin{aligned}\mathcal{A} &= [\mathcal{A}_1 \quad \mathcal{A}_2] \\ \mathcal{A}_1 &= [(C \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}] \\ \mathcal{A}_2 &= [(C \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(AL \otimes AL) + (I_p \otimes \Gamma)]\end{aligned} \tag{5.31}$$

Theorem 5.2. *If $\mathcal{A}$ has full column rank then in the limit of large data: $N_d \to \infty$, where N_d is the length of the innovations sequence used to estimate $\hat{\mathcal{R}}_1(N)$, the ALS technique gives unbiased estimates for the covariances:* $\lim_{N_d \to \infty} \hat{\mathbb{Q}} \to TG_pQ_aG_p^TT^T$ *and* $\lim_{N_d \to \infty} \hat{R}_v \to R_v$.

Proof. See Odelson et al. (2006b) for proof showing that the estimates are unbiased. □

Lemma 5.3. *If covariances $TG_pQ_aG_p^TT^T$ and R_v are used to calculate the Kalman filter for the incorrect disturbance model in Equation 5.22, then the resulting innovations sequence is the same as when using the optimal Kalman filter with the actual plant disturbance model in Equation 5.21.*

Proof. Substituting $TG_pQ_aG_p^TT^T$ and R_v in the Kalman filter Equation 5.6 for the incorrect model in Equation 5.22 and using the transformations in Equation 5.23, we get:

$$L_m = TL_p$$

Equivalence and optimality of the innovations sequence then follows from Lemma 5.2 and noting that L_m is the optimal Kalman gain for the disturbance model in Equation 5.21. □

Thus, even when the disturbance model is chosen different from the actual plant disturbance location, the ALS technique can be used to estimate transformed covariances such that the closed-loop performance is equivalent to that achieved using the optimal filter and the actual plant disturbance model.

Necessary and sufficient conditions for uniqueness of the ALS problem in Equation 5.30 are given by Theorem 3.1 in Chapter 3. Its informative to note that $\mathcal{A}$ in Equation 5.30 may not be full rank giving nonunique estimates to the ALS optimization. The nonuniqueness however, is not a limitation as shown in Theorem 5.3.

Theorem 5.3. *Given that the ALS optimization in Equation 5.30 is not unique, every nonunique solution gives the same Kalman filter gain.*

Proof. From Lemma 3.1 in Chapter 3, since the ALS optimization is not unique the matrix $\mathcal{A}$ defined in Equation 5.31 is not full column rank.

In addition from the statement and the proof to Lemma 3.2 in Chapter 3 we have the following condition:

$$(C \otimes I_n)(I_{n^2} - (\bar{A} \otimes \bar{A}))^{-1}\mathcal{D}_{n_a}(\mathbb{Q}_N)_{ss} = 0 \tag{5.32}$$

where, n_a is the dimension of $\mathbb{Q}$ and $\begin{bmatrix} (\mathbb{Q}_N)_{ss} \\ 0 \end{bmatrix}$ is an element in the null space of the $\mathcal{A}$. The subscript 'ss' denotes the column-wise stacking of only the symmetric elements of the matrix $\mathbb{Q}_N$. Define P_N uniquely as: $(P_N)_s = (I_{n^2} - \bar{A} \otimes \bar{A})^{-1}\mathcal{D}_{n_a}(\mathbb{Q}_N)_{ss}$. Notice that restacking the above vectors column-wise into a matrix and using rules of Kronecker

products gives the following Lyapunov Equation:

$$P_N = \bar{A} P_N \bar{A}^T + \mathbb{Q}_N$$

Also from Equation 5.32, we see that P_N is symmetric and satisfies

$$P_N C^T = 0, \qquad C P_N = 0 \tag{5.33}$$

giving:

$$P_N = A P_N A^T + \mathbb{Q}_N \tag{5.34}$$

Let $\hat{\mathbb{Q}}, \hat{R}_v$ be estimates from the ALS optimization. From the proof of Lemma 3.2 we have that $\hat{R}_v$ is a unique estimate and $\hat{\mathbb{Q}}$ is any arbitrary nonunique estimate. Also, every nonunique estimate of $\mathbb{Q}$ is characterized by $\hat{\mathbb{Q}} + \alpha \mathbb{Q}_N$ where α is a scalar, provided the sum is positive semidefinite.

Let the Kalman filter gain using the above estimates be $\hat{L}$ calculated as:

$$\begin{aligned} \hat{P} &= A_m \hat{P} A_m^T - A_m \hat{P} C_m^T (C_m \hat{P} C_m^T + \hat{R}_v)^{-1} C_m \hat{P} A_m^T + \hat{\mathbb{Q}} \\ \hat{L} &= \hat{P} C_m^T (C_m \hat{P} C_m^T + \hat{R}_v)^{-1} \end{aligned} \tag{5.35}$$

If we define $\tilde{P} = \hat{P} + \alpha P_N$ and $\tilde{\mathbb{Q}} = \hat{\mathbb{Q}} + \alpha \mathbb{Q}_N$, then using Equations 5.33 and 5.34, we see that $\tilde{P}$ and $\tilde{\mathbb{Q}}$ can be substituted in Equation 5.35 to produce the same Kalman filter gain $\hat{L}$. Thus, the statement in this theorem is proved. □

5.5 Simulation Examples

The following examples illustrate the ideas developed in the previous sections. Closed-loop performance of a pure input disturbance model ($B_d = B, C_d = 0$) is compared

against the pure output disturbance model ($B_d = 0, C_d = I_p$) for the following MIMO system ($n_d = p$ in both models):

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{s+1} & \frac{2}{3s+1} \\ 0 & \frac{2s+1}{(3s+1)(s+2)} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

A minimal state-space realization (discrete-time model with a sampling time of 1 sec) gives:

$$A = \begin{bmatrix} 0.368 & 0 & 0 \\ 0 & 0.455 & 0.050 \\ 0 & 0.453 & 0.629 \end{bmatrix}, \quad B = \begin{bmatrix} -0.632 & 0 \\ 0 & -0.971 \\ 0 & -0.401 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & -0.471 & -0.272 \\ 0 & -0.471 & 0 \end{bmatrix}$$

5.5.1 Deterministic Disturbances

In the plant, deterministic unmodelled disturbances are added to both the inputs as shown in Figure 5.4. The disturbances entering are of different magnitude and are uncorrelated. Other stochastic disturbances are absent in the simulation: $Q_w, R_v = 0$. An input disturbance model is an accurate description of the unmodelled disturbance in the plant. The optimal state estimator is the deadbeat estimator with the gain calculated by using $G = I, Q_w = 0, R_v \approx 0, Q_\xi \neq 0$ in the estimator Riccati Equation 5.6. The

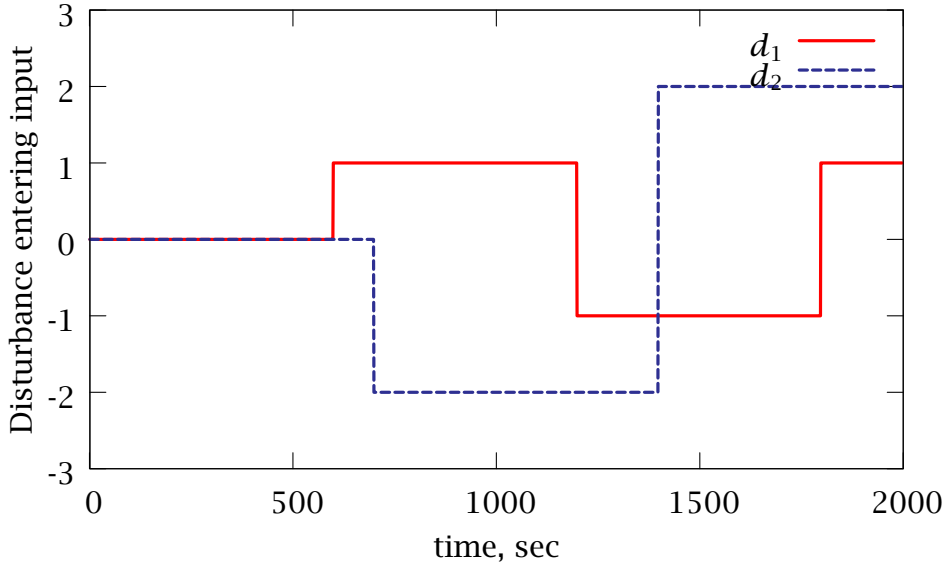


Figure 5.4: Disturbances entering the inputs

deadbeat estimator for the input disturbance model is:

$$L_1 = \begin{bmatrix} -1 & 1.24 \\ 0 & -2.12 \\ 0 & -0.88 \\ 1.58 & -1.96 \\ 0 & 2.18 \end{bmatrix}$$

If on the other hand the disturbance is incorrectly modelled as entering entirely in the output (the output disturbance model), the same covariances $G = I, Q_w = 0, R_v \approx$

$0, Q_\xi \neq 0$ give the following deadbeat estimator gain:

$$L_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The controller penalties used in Equation 5.8 were $Q = I, R = I$. The initial data was simulated with the output disturbance model (incorrect) with the above initial suboptimal estimator gain. The data was then processed using the ALS technique from Section 5.4 to estimate (transformed) covariances GQ_aG^T, R_v and then used to calculate the estimator gain. The estimator gains from the different cases is presented in Table 5.1. The optimum gain is the deadbeat estimator with the input disturbance model. As seen in Table 5.1, using the ALS technique the estimator gains are close the optimum values, even when the incorrect disturbance model is used.

A snapshot of the performance of the closed-loop with the different models and estimator gains in shown in Figures 5.5.

As seen in the Figures 5.5, once the estimator gain is found using covariances estimated using the ALS technique, the inputs and the outputs become equivalent and thus indistinguishable using either of the disturbance models.

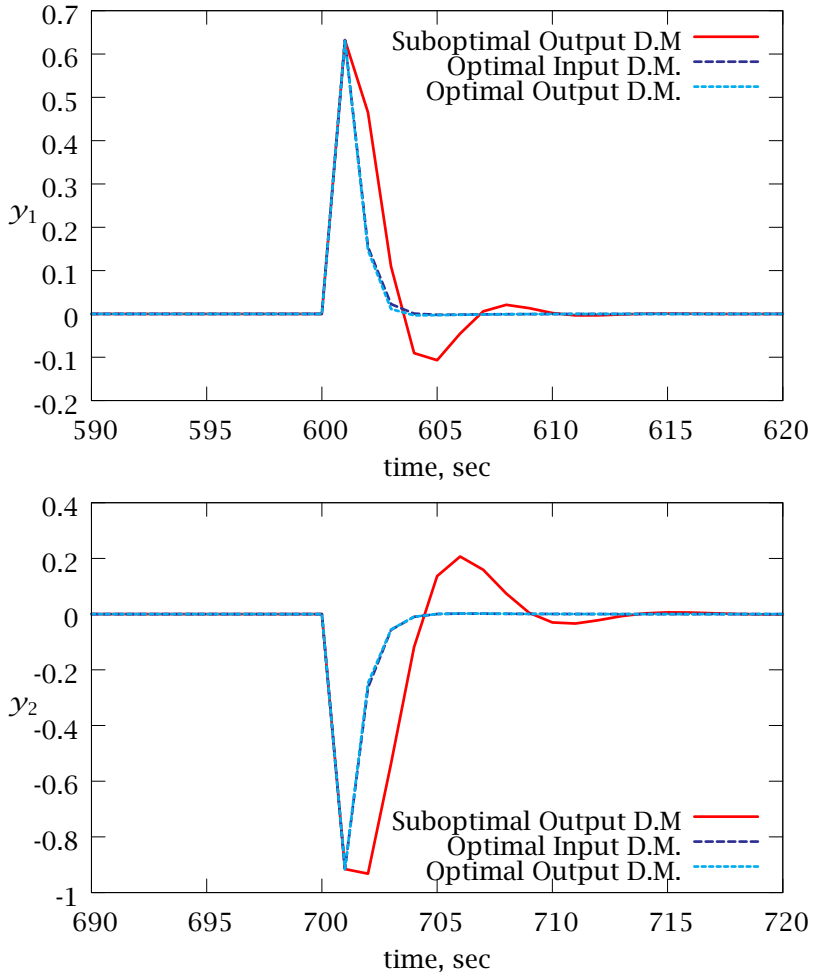
The closed-loop cost benefits for the deterministic disturbances shown in Figure 5.4 can be easily calculated and is presented in Table 5.2.

	Deadbeat Estimator	ALS Estimator	Transformation
Input D.M.	$L_{opt} = \begin{bmatrix} -1 & 1.24 \\ 0 & -2.12 \\ 0 & -0.88 \\ 1.58 & -1.96 \\ 0 & 2.18 \end{bmatrix}$	$\hat{L}_{inp} = \begin{bmatrix} -0.891 & 1.11 \\ -0.08 & -1.95 \\ -0.05 & -0.93 \\ 1.44 & -1.73 \\ 0.1 & 2.00 \end{bmatrix}$	$T^{-1}\hat{L}_{out} = \begin{bmatrix} -0.88 & 1.01 \\ -0.06 & -2.03 \\ -0.04 & -0.61 \\ 1.45 & -1.76 \\ 0.1 & 2.04 \end{bmatrix}$
Output D.M.	$L_{ini} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\hat{L}_{out} = \begin{bmatrix} 0.59 & -0.75 \\ 0.16 & 2.30 \\ 0.34 & 6.88 \\ 1.68 & 2.31 \\ 0.1 & 2.04 \end{bmatrix}$	$T\hat{L}_{inp} = \begin{bmatrix} 0.55 & -0.62 \\ 0.12 & 2.29 \\ 0.31 & 6.42 \\ 1.64 & 2.27 \\ 0.1 & 2.00 \end{bmatrix}$

Table 5.1: Table illustrating equivalence in the estimator gains identified from data

5.5.2 Stochastic Disturbances

Instead of the deterministic disturbance, in this section, we consider slow moving integrated white noise disturbances enter each input as shown in Figure 5.6. The actual plant disturbance model in this case is again the input disturbance model and the out-



put disturbance model performs poorly without the covariance estimates using the ALS technique. The covariance of the integrating disturbances is $Q_\xi = 10^{-3}I$. In addition the inputs and the outputs are corrupted by zero mean Gaussian white noise having covariances $Q_w = 10^{-4}I, R_v = 10^{-5}I$.

The estimator gains for the input and the output disturbance models were calculated separately using the ALS technique (Equation 5.30) from the same data set and a snapshot of the performance of the two disturbance models is shown in Figure 5.7. Again the inputs and outputs perform very similarly irrespective of the disturbance

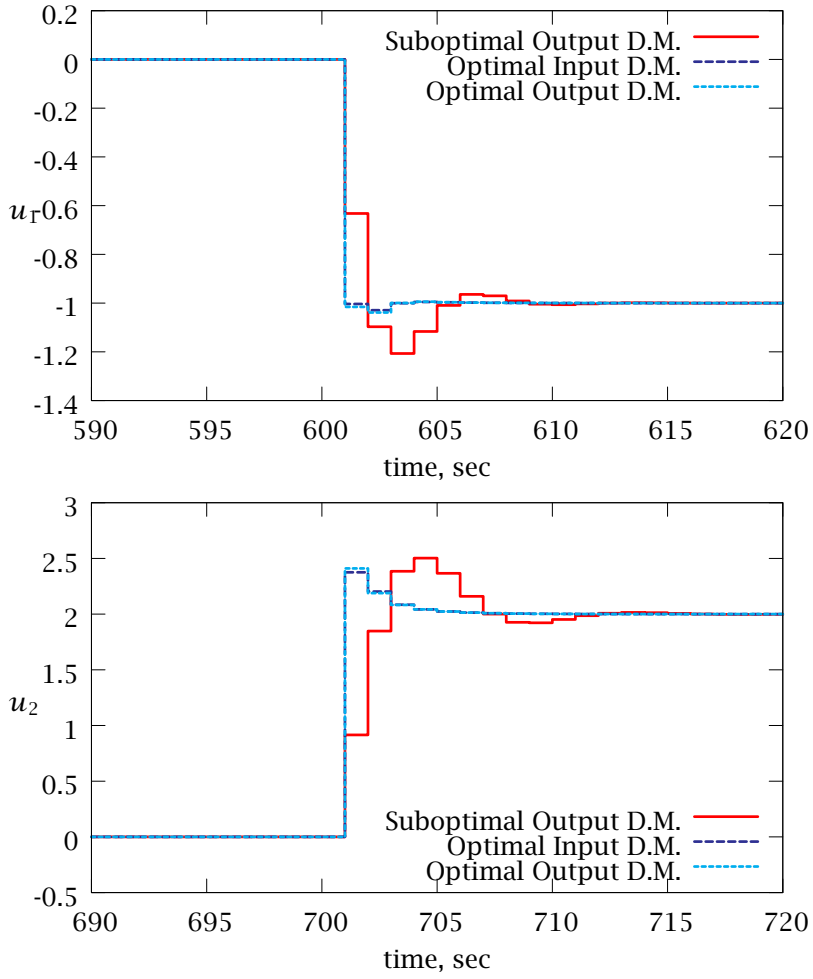


Figure 5.5: Closed-loop performance using estimator gains from the ALS technique (D.M.=Disturbance Model)

model chosen once the ALS technique is used to calculate the estimator gains.

The comparison between the expected closed-loop costs (see Appendix 5.7.2 for calculation of closed-loop cost) for the three cases is presented in Table 5.3. Note that the optimal closed-loop cost is 1.788×10^{-4}.

		Initial	ALS	Benefits
Input	$\sum x^T Q x$	N/A	15.76	N/A
Disturbance	$\sum (u-u_s)^T R (u-u_s)$	N/A	1.82	
Model	Total	N/A	17.58	
Output	$\sum x^T Q x$	34.48	15.60	~ 49%
Disturbance	$\sum (u-u_s)^T R (u-u_s)$	0	1.69	
Model	Total	34.48	17.30	

Table 5.2: Table showing closed-loop benefits when using the ALS technique

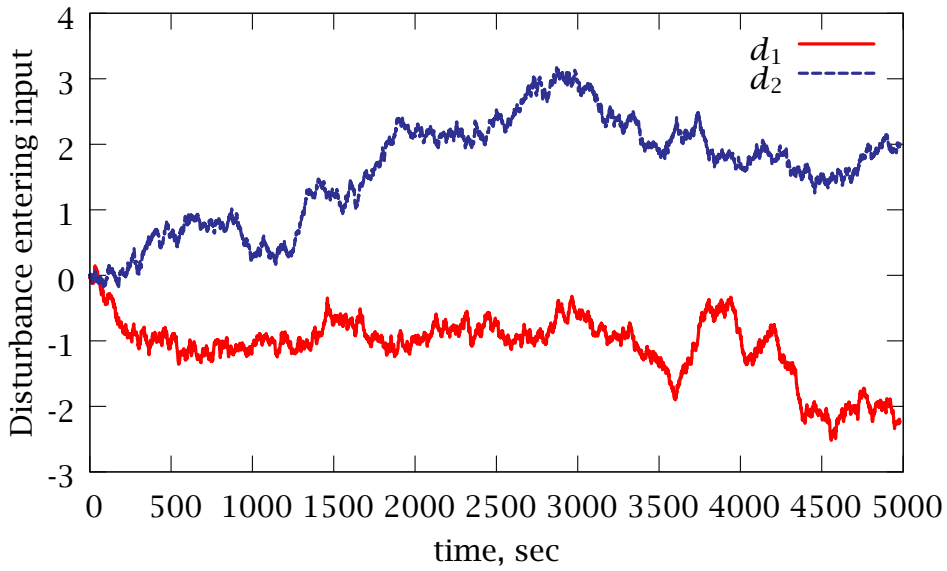
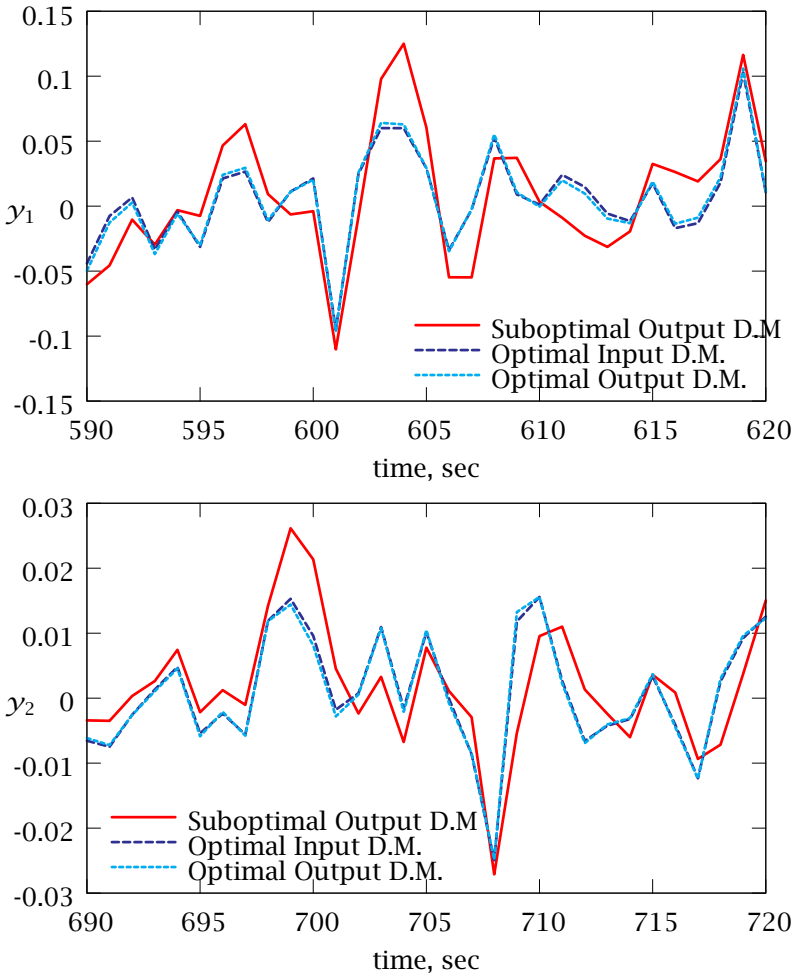


Figure 5.6: Disturbances entering the inputs



	Stochastic Closed-loop Cost	Initial ($\times 10^4$)	ALS ($\times 10^4$)	Benefits
Input D.M.	$E[x_k^T Q x_k + u_k^T R u_k]$	N/A	1.817	N/A
Output D.M.	$E[x_k^T Q x_k + u_k^T R u_k]$	5.659	1.897	66.5%

Table 5.3: Table showing expectation of closed-loop cost for different estimator gains (Optimal Cost is 1.788×10^{-4}). See Appendix 5.7.2.

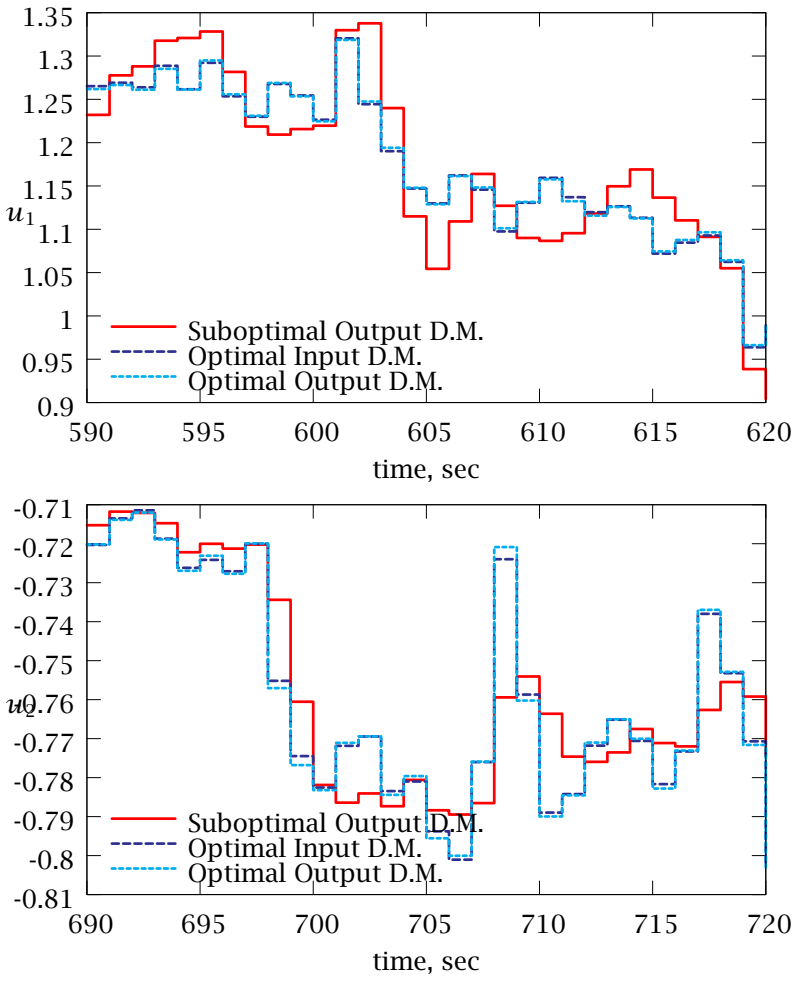


Figure 5.7: Closed-loop performance of the estimators with gains calculated using the ALS technique (D.M.=Disturbance Model)

5.6 Conclusions

Although the choice of the disturbance model for offset-free model predictive control has been the subject of active research, the choice of an appropriate estimator gain for the disturbance model has not been studied until recently (Pannocchia and Bemporad, 2007). The focus of most of the research has been to design appropriate distur-

bance models to accurately represent the actual plant disturbance source while assuming that the estimator gain is known. In Section 5.3 it was shown that since different augmented disturbance models are nonminimal realizations of the same input-output process, there exists estimator gain matrices for each disturbance model that achieve the same closed-loop input-output performance.

In Section 5.4 an autocovariance least-squares (ALS) technique was presented to estimate the covariances of the disturbance model from steady state data (closed-loop or open-loop) irrespective of the true disturbance source. The estimation of these covariances and hence the estimator gain was shown to compensate automatically for the incorrect choice of the model's disturbance location. Examples were presented to show the benefits of using the ALS technique to estimate the covariances and to illustrate the equivalent closed-loop performance using different disturbance models.

A direct implication of the results presented in this chapter is that an industrial practitioner can model the disturbance as an integrator at any location in the plant but then use the autocovariance of the data to estimate the disturbance covariances to calculate the estimator gain. The control performance using this rough disturbance model and data-based estimator gain is often comparable to that obtained using a disturbance model that accurately represents the true plant disturbance. Modeling effort can then be directed to obtaining the estimator gain from data rather than trying to model accurately the dynamics of the unknown disturbance, which is a much more complex modeling task.

5.7 Appendix

5.7.1 Proof of Lemma 5.1

Since the pair (A, C) is assumed observable, we have rank $\begin{bmatrix} A - I \\ C \end{bmatrix} = n$. From the detectability conditions in Equation 5.9 we then get:

$$\text{rank} \begin{bmatrix} B_{d1} \\ C_{d1} \end{bmatrix} = n_d, \quad \text{rank} \begin{bmatrix} B_{d2} \\ C_{d2} \end{bmatrix} = n_d \tag{5.36}$$

Substituting T partitioned as $T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$ into the equivalence relations in Equation 5.12 we get:

1. Equivalence between the augmented input matrices:

$$TB_a = B_a$$

$$\implies T_{11}B = B, \quad T_{21}B = 0$$

Since T is an invertible transformation matrix, the diagonal block T_{11} is also invertible. This implies:

$$T_{11} = I \tag{5.37}$$

2. Equivalence between the augmented state matrices:

$$TA_1 = A_2T$$

$$\implies \begin{bmatrix} A - I & B_{d2} \end{bmatrix} \begin{bmatrix} T_{12} \\ \\ T_{22} \end{bmatrix} = B_{d1}, \quad B_{d2} T_{21} = 0 \tag{5.38}$$

3. Equivalence between the augmented output matrices:

$$C_1 = C_2 T$$

$$\implies \begin{bmatrix} C & C_{d2} \end{bmatrix} \begin{bmatrix} T_{12} \\ \\ T_{22} \end{bmatrix} = C_{d1}, \quad C_{d2} T_{21} = 0 \tag{5.39}$$

Using Equations 5.38 and 5.39 we get:

$$\begin{bmatrix} A - I & B_{d2} \\ \\ C & C_{d2} \end{bmatrix} \begin{bmatrix} T_{12} \\ \\ T_{22} \end{bmatrix} = \begin{bmatrix} B_{d1} \\ \\ C_{d1} \end{bmatrix}, \qquad \begin{bmatrix} B_{d2} \\ \\ C_{d2} \end{bmatrix} T_{21} = 0 \tag{5.40}$$

Using the rank condition in Equation 5.36 in the above equation implies:

$$T_{21} = 0 \tag{5.41}$$

Putting Equations 5.37, 5.40 and 5.41 together, we get:

$$T_{11} = I, \qquad \begin{bmatrix} A - I & B_{d2} \\ \\ C & C_{d2} \end{bmatrix} \begin{bmatrix} T_{12} \\ \\ T_{22} \end{bmatrix} = \begin{bmatrix} B_{d1} \\ \\ C_{d1} \end{bmatrix}, \qquad T_{21} = 0$$

□

5.7.2 Closed-Loop Cost

Let the model of the plant be given by:

$$x_{k+1} = Ax_k + Bu_k + Gw_k \tag{5.42}$$

$$y_k = Cx_k + v_k \tag{5.43}$$

where, w_k and v_k are the noises corrupting the state and the measurements and having covariances Q_w and R_v respectively.

A linear estimator (not necessarily optimal) for the state has the following form:

$$\hat{x}_{k|k-1} = A\hat{x}_{k|k} + Bu_k \tag{5.44}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L(y_k - C\hat{x}_{k|k-1}) \tag{5.45}$$

where, $\hat{x}_{k|k}$ is the filtered estimate given the measurements and inputs up to time t_k and $\hat{x}_{k|k-1}$ is the predicted estimate given measurements and inputs up to time t_{k-1}.

The deterministic LQR objective is given by:

$$\Phi = \sum_{k=0}^{\infty} x_k^T Q x_k + u_k^T R u_k$$

giving the feedback law $u_k = K\hat{x}_{k|k}$. For the stochastic case, the objective can be defined as:

$$\Phi = \lim_{N_c \to \infty} \frac{1}{N_c} E\left[\sum_{k=0}^{N_c} x_k^T Q x_k + u_k^T R u_k\right] \tag{5.46}$$

The closed-loop cost for the continuous time state-space model is derived in Anderson and Moore (1990, p. 220): Simple algebraic manipulations on Equations 5.42 and 5.44 give the following augmented system:

$$\begin{bmatrix} x_{k+1} \\ \hat{x}_{k+1|k+1} - x_{k+1} \end{bmatrix} = \underbrace{\begin{bmatrix} A + BK & BK \\ 0 & A - LCA \end{bmatrix}}_{\tilde{A}} \begin{bmatrix} x_k \\ \hat{x}_{k|k} - x_k \end{bmatrix} + \underbrace{\begin{bmatrix} G & 0 \\ (LC - I)G & L \end{bmatrix}}_{\tilde{G}} \begin{bmatrix} w_k \\ v_k \end{bmatrix} \tag{5.47}$$

The covariance matrix for the augmented system:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{bmatrix} = E\left[\begin{bmatrix} x_k \\ \hat{x}_{k|k} - x_k \end{bmatrix}\begin{bmatrix} x_k^T & \hat{x}_{k|k}^T - x_k^T \end{bmatrix}\right] \tag{5.48}$$

S is given by the following Lyapunov equation:

$$S = \tilde{A}S\tilde{A}^T + \tilde{G}\begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}\tilde{G}^T \tag{5.49}$$

The closed-loop cost in Equation 5.46 can now be written as (using rules for trace and expectations):

$$\begin{aligned}
\Phi &= \lim_{N_c\to\infty}\frac{1}{N_c}E\left[\sum_{k=0}^{N_c} x_k^T Q x_k + u_k^T R u_k\right] \\
&= \lim_{N_c\to\infty}\frac{1}{N_c}\sum_{k=0}^{N_c}\mathrm{tr}\left[QE(x_k x_k^T) + K^T RKE(\hat{x}_{k|k}\hat{x}_{k|k}^T)\right] \\
&= \lim_{N_c\to\infty}\frac{1}{N_c}\sum_{k=0}^{N_c}\mathrm{tr}\left[QS_{11} + K^T RK(S_{11} + S_{12} + S_{12}^T + S_{22})\right] \\
&= \mathrm{tr}\left[QS_{11} + K^T RK(S_{11} + S_{12} + S_{12}^T + S_{22})\right]
\end{aligned}$$

5.7.3 Alternate Derivation of the Transformation matrix

To find the transformation matrix, T in Equation 5.12 we use the following rules of Kronecker products:

$$\text{If, } AXC^T = B$$

$$\begin{bmatrix} b_{11} \\ b_{21} \\ b_{12} \\ b_{22} \end{bmatrix} = (C \otimes A) \begin{bmatrix} x_{11} \\ x_{21} \\ x_{12} \\ x_{22} \end{bmatrix}$$

$$\text{or, } B_s = (C \otimes A)X_s$$

where, the subscript '*s*' indicates that the elements have been stacked.

1.

$$TA_1T^{-1} = A_2$$

$$\therefore TA_1 = A_2T$$

$$\therefore [(A_1^T \otimes I_{n+n_d}) - (I_{n+n_d} \otimes A_2)]T_s = 0_{(n+n_d)^2,1}$$

2.

$$TB_a = B_a$$

$$\therefore (B_a^T \otimes I_{n+n_d})T_s = [B_a]_s$$

3.

$$C_1 = C_2T$$

$$\therefore (I_{n+n_d} \otimes C_2)T_s = [C_1]_s$$

Putting all of the above together, we get:

$$\underbrace{\begin{bmatrix} B_a^T \otimes I_{n+n_d} \\ I_{n+n_d} \otimes C_2 \\ (A_1^T \otimes I_{n+n_d}) - (I_{n+n_d} \otimes A_2) \end{bmatrix}}_{\tilde{A}} T_s = \underbrace{\begin{bmatrix} (B_a)_s \\ (C_1)_s \\ 0_{(n+n_d)^2,1} \end{bmatrix}}_{\tilde{B}}$$

T_s has a unique solution if:

$$\text{rank}([\tilde{A}, \tilde{B}]) = (n + n_d)^2$$

i.e. $\tilde{B}$ has to be in the range of $\tilde{A}$.

Chapter 6

Connections between ALS and Maximum Likelihood Estimation (MLE)

It is well known that a Bayesian estimation procedure gives the minimum variance estimates for the parameters when the errors in the measurements are distributed normally (Box and Tiao (1973, p. 310);Anderson and Moore (1979, p. 32)). In the ALS technique, the estimated parameters are the covariances Q_w, R_v. However, the errors in the parameter estimates are not distributed normally. Section 4.1 in Chapter 4 presents more details about the minimum variance estimation procedure for the ALS technique. In this chapter, we formulate the maximum likelihood estimation (MLE) procedure for the covariances Q_w, R_v given a set of data $\{y_1, \cdots, y_{N_d}\}$ and $\{u_1, \cdots, u_{N_d}\}$ for the linear time-invariant state-space model. We also find the connections between the ALS formulation and the MLE formulation and present an iterative technique to find the maximum likelihood estimates using the ALS formulation.

The Bayesian estimator for the covariances Q_w, R_v involves finding the posterior expectation of the covariances conditional on the given data. In the limit of large data or an uninformative prior, the *à-priori* distribution of the covariances becomes negligible and Bayesian estimation becomes equivalent to the maximum likelihood estimation. ML estimation procedures aim to maximize the likelihood function with respect to the parameters.

The literature on solving the MLE problem usually include the state-space matrices A, B, C as unknowns in addition the covariances Q_w, R_v e.g Schweppe (1965). If we start with the linear time-invariant state-space model and subtract out the state estimates from the actual state we get the following L-innovations form of the model as shown in Equations 3.1a and 3.1b in Chapter 3:

$$\varepsilon_{k+1} = \underbrace{(A - ALC)}_{\bar{A}} \varepsilon_k + \underbrace{\begin{bmatrix} G & -AL \end{bmatrix}}_{\bar{G}} \begin{bmatrix} w_k \\ v_k \end{bmatrix} \tag{6.1a}$$

$$y_k = C\varepsilon_k + v_k \tag{6.1b}$$

The unknown noise covariance is:

$$\operatorname{cov}\left(\begin{bmatrix} w_k \\ v_k \end{bmatrix} \right) = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}$$

We can then write the innovations form of the log likelihood function as shown in Schweppe (1965):

$$-2 \ln L_Y(\theta) = \sum_{k=1}^{N_d} \ln|\Sigma(\theta)| + \sum_{k=1}^{N_d} y_k(\theta)^T \Sigma(\theta)^{-1} y_k(\theta) \tag{6.2}$$

in which $L_Y(\theta)$ is the likelihood function and $\Sigma = CPC^T + R_v$ with P the covariance of ε_k. θ is the vector containing the unknowns A, B, C, G, Q_w, R_v. If the state-space matrices A, B, C, G are assumed to be known, then the estimation of the unknown covariances Q_w, R_v follows an iterative scheme:

- Guess Q_w, R_v and stack them in a vector $\theta = [(Q_w)_s, (R_v)_s]$
- Calculate $\Sigma = CPC^T + R_v$ and $\mathcal{Y}_k$ using the guessed values for the covariances
- Compute the innovations form of the likelihood function from Equation 6.2 and find a gradient direction to maximize $L_Y(\theta)$
- Iterate until the estimates converge

As emphasized by Shumway and Stoffer (2000, chap. 4), the L-innovations $\mathcal{Y}_k$ are dependent on the parameter θ and the likelihood function is a highly nonlinear and complicated function of θ. Newton-Raphson methods have been suggested in the literature to recursively update the parameters and to maximize the objective function in Equation 6.2. A more robust algorithm was given by Shumway and Stoffer (1982) based on the EM (expectation-maximization) algorithm originally used by Dempster et al. (1977). The EM algorithm is also an iterative procedure that like other algorithms, fails to converge to the global optimum starting from a bad initial guess.

In this chapter, we present a simplification of the maximum likelihood estimation technique using matrix derivatives and the ALS formulation in Chapter 3.

6.1 Maximum Likelihood Formulation

Let the set of L-innovations data from Equation 6.1a and 6.1b be obtained when the plant is operating at steady state, thus not needing any exciting inputs to be applied to the plant. The L-innovations are then independent of the input sequence and do not depend on the controller. Let the history of discrete measurements for the times $\{t_1, \cdots t_{N_d}\}$ be available: $\{\mathcal{Y}_1, \cdots \mathcal{Y}_{N_d}\}$. We wish to find the maximum likelihood estimates for the covariances Q_w, R_v given the model in Equations 6.1a and 6.1b.

Define $Y_p = [\mathcal{Y}_1^T, \cdots \mathcal{Y}_{N_d}^T]^T$ to be the set of L-innovations. Then the distribution of Y_p is given by (see Chapter 4, Appendix 4.4.1 for derivation):

$$Y_p \sim N(0, \Omega_p)$$

where,

$$\begin{aligned}\Omega_p = &\underbrace{\begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N_d-1} \end{bmatrix}}_{\mathcal{O}} P\mathcal{O}^T + \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} + \Psi \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} \\ &+ \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ C\bar{G} & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots \\ C\bar{A}^{N_d-2}\bar{G} & \cdots & C\bar{G} & 0 \end{bmatrix}}_{\Gamma_f} \begin{bmatrix} \bar{Q}_w & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{Q}_w \end{bmatrix} \Gamma_f^T + \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} \Psi^T\end{aligned} \tag{6.3}$$

and,

$$\Psi = \Gamma_f \begin{bmatrix} -AL & 0 & 0 & 0 \\ 0 & -AL & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & -AL \end{bmatrix}, \qquad \bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}$$

In Kronecker product notation, we can write the following:

$$\Omega_p = \mathcal{O}P\mathcal{O}^T + \Gamma \bigoplus_{i=1}^{N_d} (\bar{G}\bar{Q}_w\bar{G}^T)\Gamma^T + \Psi \bigoplus_{i=1}^{N_d} R_v + \bigoplus_{i=1}^{N_d} R_v\Psi^T + \bigoplus_{i=1}^{N_d} R_v \tag{6.4}$$

and P is given by the Lyapunov equation:

$$P = \underbrace{(A - ALC)}_{\bar{A}} P (A - ALC)^T + \underbrace{\begin{bmatrix} G & -AL \end{bmatrix}}_{\bar{G}} \underbrace{\begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}}_{\bar{Q}_w} \begin{bmatrix} G^T \\ -L^T A^T \end{bmatrix} \tag{6.5}$$

The likelihood function $L(Q_w, R_v|Y_p)$ is given by writing the probability:

$$p(y_1, \cdots y_{N_d}|Q_w, R_v)$$

as a function of Q_w, R_v. Since the distribution Y_p is normal, we can write the conditional probability density function as:

$$L(Q_w, R_v|Y_p) = p(Y_p|Q_w, R_v) = \frac{1}{(2\pi)^{\frac{pN_d}{2}} |\Omega_p|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \begin{bmatrix} y_1^T & \cdots y_{N_d}^T \end{bmatrix} \Omega_p^{-1} \begin{bmatrix} y_1 \\ \vdots \\ y_{N_d} \end{bmatrix} \right) \tag{6.6}$$

The maximum likelihood estimate is given by maximizing the log of the likelihood function (or minimizing the negative log of the likelihood) with the covariances Q_w, R_v as the parameters, given the data Y_p.

$$\min_{Q_w,R_v} -2\ln L(Q_w,R_v|Y_p)$$

$$\text{or,} \quad \min_{Q_w,R_v} pN_d\ln(2\pi) + \ln|\Omega_p| + Y_p^T\Omega_p^{-1}Y_p$$

To get an analytical formula for the minimizer of the maximum likelihood function, we set the derivative of $-\ln L(Q_w,R_v|Y_p)$ with respect to Q_w, R_v to zero. Since the quantities in the above equation are scalars, we can take their Trace. The total derivative for the above log likelihood function can be written as (Magnus and Neudecker, 1999, chap. 9) (since $pN_d\ln(2\pi)$ is a constant):

$$\begin{aligned}
-2d\ln L &= \text{Tr}\,(\Omega_p^{-1}d\Omega_p) + \text{Tr}\,(-\Omega_p^{-1}d\Omega_p\Omega_p^{-1}Y_pY_p^T) \\
&= \text{Tr}\,(\Omega_p^{-1}d\Omega_p(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T)) \\
&= \mathbb{E}_{pN_d}[(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T)\otimes\Omega_p^{-1}](d\Omega_p)_s
\end{aligned}$$

where, $\mathbb{E}_{\cdot}$ is the single row matrix with 1's and 0's to add up the diagonals of the original matrix. For example:

$$\text{Tr}\left(\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix}\right) = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbb{E}_3} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_9 \end{bmatrix}}_{A_s}$$

The notation with subscript 's' is used to denote the column-wise stacking of the matrix. Writing the total derivatives as above facilitates taking the derivatives with respect to the covariances and deriving the necessary conditions on the maximum likelihood estimates.

Using Equation 6.4 to write $d\Omega$ in terms of dQ_w and dR_v, we get:

$$-2d\ln L = \mathbb{E}_{pN_d}[(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T) \otimes \Omega_p^{-1}][\mathcal{O}dP\mathcal{O}^T + \Gamma_f \bigoplus_{i=1}^{N_d} dQ_w \Gamma_f^T + \Psi \bigoplus_{i=1}^{N_d} dR_v + \bigoplus_{i=1}^{N_d} dR_v \Psi^T + \bigoplus_{i=1}^{N_d} dR_v]_s$$

Using rules of Kronecker products and Equation 6.5 to write the estimate error covariance P_s in terms of Q_w, R_v and collecting terms in $(dQ_w)_s$ and $(dR_v)_s$, we get:

$$\begin{aligned} -2d\ln L &= \mathbb{E}_{pN_d}[(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T) \otimes \Omega_p^{-1}]D(G \otimes G)(dQ_w)_s \\ &\quad + \mathbb{E}_{pN_d}[(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T) \otimes \Omega_p^{-1}][D(AL \otimes AL) + (\Psi \oplus \Psi + I_{p^2N_d^2})\mathcal{I}_{p,N_d}](dR_v)_s \\ \therefore -2d\ln L &= \mathbb{A}_1(dQ_w)_s + \mathbb{A}_2(dR_v)_s \end{aligned} \tag{6.7}$$

where,

$$\begin{aligned} D &= (\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma)\mathcal{I}_{n,N_d} \\ \mathbb{A}_1 &= \mathbb{E}_{pN_d}[(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T) \otimes \Omega_p^{-1}]D(G \otimes G) \\ \mathbb{A}_2 &= \mathbb{E}_{pN_d}[(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T) \otimes \Omega_p^{-1}][D(AL \otimes AL) + (\Psi \oplus \Psi + I_{p^2N_d^2})\mathcal{I}_{p,N_d}] \end{aligned}$$

and $\mathcal{I}_{p,N}$ is the permutation matrix containing 1's and 0's that gives the relation:

$$\left(\bigoplus_{i=1}^{N} R_v\right)_s = \mathcal{I}_{p,N}(R_v)_s$$

From Equation 6.7 it follows that the derivatives of scalar function $-2d\ln L$ with respect to the matrices Q_w and R_v are given by:

$$\frac{-2d\ln L}{dQ_w} = [\mathbb{A}_1]_{g\times g}$$
$$\frac{-2d\ln L}{dR_v} = [\mathbb{A}_2]_{p\times p}$$

where $[A_1]_{g,g}$ represents creating a $\mathbb{R}^{g\times g}$ matrix from the g^2 elements in the vector A_1 by accessing the elements column-wise.

Setting the derivatives to zero, we thus have the necessary conditions for the solution to the maximum likelihood estimation problem:

$$\frac{-2d\ln L}{dQ_w} = 0, \quad \Longrightarrow [\mathbb{A}_1]_{g\times g} = 0$$
$$\frac{-2d\ln L}{dR_v} = 0, \quad \Longrightarrow [\mathbb{A}_2]_{p\times p} = 0$$

The above conditions are expressed equivalently as:

$$\mathbb{E}_{pN_d}[(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T)\otimes\Omega_p^{-1}]\begin{bmatrix}\mathcal{A}_1 & \mathcal{A}_2\end{bmatrix} = 0$$

where we have:

$$\begin{aligned}\mathcal{A}_1 &= [(\mathcal{O}\otimes\mathcal{O})(I_{n^2} - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N_d}](G\otimes G)\\ \mathcal{A}_2 &= [(\mathcal{O}\otimes\mathcal{O})(I_{n^2} - \bar{A}\otimes\bar{A})^{-1} + (\Gamma\otimes\Gamma)\mathcal{I}_{n,N_d}](AL\otimes AL)\\ &\quad + (\Psi\oplus\Psi + I_{p^2N^2})\mathcal{I}_{p,N_d}\end{aligned} \tag{6.8}$$

To simplify these equations further, we note the following property of Kronecker products. If A is a $\mathbb{R}^{n\times n}$ matrix then we have:

$$[\mathbb{E}_n(I_n\otimes A)]_{n\times n} = A^T$$

Using this property and other properties of Kronecker products (Magnus and Neudecker (1999, chap. 2);Graham (1981, chap. 2);Steeb (1991)), we get the following simplification to the necessary conditions listed above:

$$[(\Omega_p^{-1})_s^T - (\Omega_p^{-1})_s^T(\Omega_p^{-1}Y_pY_p^T \otimes I_{pN_d})] \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 \end{bmatrix} = 0$$

Thus, the algebraic equations obtained by setting the derivatives of $-2\ln L$ w.r.t. Q_w, R_v to 0 are:

$$[\Omega_p^{-1}(I_{pN_d} - \Omega_p^{-1}Y_pY_p^T)]_s^T \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 \end{bmatrix} = 0$$

These conditions can also be written as:

$$\mathcal{A}^T(\Omega_p^{-1})_s - \mathcal{A}^T(I_{pN_d} \otimes \Omega_p^{-2})(Y_pY_p^T)_s = 0 \tag{6.9}$$

where, $\mathcal{A} = \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 \end{bmatrix}$ and $\mathcal{A}_1, \mathcal{A}_2$ are defined in Equation 6.8. Thus, the maximum likelihood estimates for Q_w, R_v are obtained by solving the above necessary conditions, where Ω_p is a nonlinear function of the parameters Q_w, R_v given by Equation 6.4.

6.1.1 Connections of MLE and the ALS formulation when $N = N_d$

We next compare the nonlinear algebraic equation for the MLE necessary conditions in Equation 6.9 with the ALS formulation to find Q_w, R_v. Given the results in Chapter 3, if the ALS problem is solved with the window size chosen as $N = N_d$, then the ALS

optimization in Equation 3.36 is written as:

$$\min_{Q_w,R_v} \left\| \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 \end{bmatrix} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \left(Y_p Y_p^T\right)_s \right\|_{W^{-1}}^2 \tag{6.10}$$

$$\text{subject to}, \quad Q_w, R_v \geq 0$$

Ignoring the constraints for the moment, the necessary conditions for the solution of the above ALS optimization are obtained by setting the derivative of the ALS objective w.r.t. Q_w, R_v to zero. The derivatives are given by:

$$\mathcal{A}^T W^{-1} \mathcal{A} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \mathcal{A}^T W^{-1} (Y_p Y_p^T)_s = 0 \tag{6.11}$$

where, $\mathcal{A} = \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 \end{bmatrix}$.

Comparing the MLE necessary conditions in Equation 6.9 with the ALS necessary conditions in Equation 6.11, we seek to find the W matrix that makes the two equations equivalent. Noting that $\mathcal{A} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} = (\Omega_p)_s$, Equation 6.11 simplifies to:

$$\mathcal{A}^T W^{-1} (\Omega_p)_s - \mathcal{A}^T W^{-1} (Y_p Y_p^T)_s = 0$$

From rules of Kronecker products we have $(I_{pN_d} \otimes \Omega_p^{-2})(\Omega_p)_s = (\Omega_p^{-1})_s$. We then find that the following choice for the W matrix makes Equations 6.9 and 6.11 equivalent:

$$W^{-1} = (I_{pN_d} \otimes \Omega_p^{-2}) \tag{6.12}$$

Thus, to solve the MLE problem as a least-squares problem, we need to set $N = N_d$ and the least-squares weight as given in Equation 6.12. However, note that the weight W is still a complicated nonlinear function of the unknown covariances Q_w, R_v and an iterative procedure is required to solve Equation 6.11.

An iterative procedure to calculate the maximum likelihood estimates using least-squares is:

1. Guess a value for the covariances: $\hat{Q}_w, \hat{R}_v$ and calculate $\hat{\Omega}_p$ from Equation 6.4

2. Calculate the ML weighting $\hat{W}^{-1}$ using Equation 6.12

3. Use the weighted ALS formulation with the weight $\hat{W}^{-1}$ to estimate new set of covariances $\hat{Q}_w, \hat{R}_v$ using the optimization in Equation 6.10

4. Repeat the above steps till convergence

6.1.2 MLE and ALS with Window Size N

In the previous section the comparison between the ALS and the MLE formulations was obtained by choosing the ALS window size as $N = N_d$. When N is chosen to be smaller than the data length N_d, similarities can still be seen between the ALS and MLE formulations.

As in the previous section, we set the derivative of the ALS objective with a window

size N to zero to obtain the necessary conditions for the minima.

$$\tilde{\mathcal{A}}^T W^{-1} \tilde{\mathcal{A}} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \tilde{\mathcal{A}}^T W^{-1} (\mathbb{Y}\mathbb{Y}^T)_s = 0 \tag{6.13}$$

where, $\mathbb{Y}$ is defined as:

$$\mathbb{Y} \triangleq \begin{bmatrix} y_1 & y_2 & \cdots & y_{N_d-N+1} \\ y_2 & y_3 & \cdots & y_{N_d-N+2} \\ \vdots & \vdots & \vdots & \vdots \\ y_N & y_{N+1} & \vdots & y_{N_d} \end{bmatrix} \tag{6.14}$$

and $\tilde{\mathcal{A}}$ is defined as in Equation 6.8 but with N instead of N_d.

We recover the covariance of $(\mathbb{Y})_s$ from Ω_p defined in Equation 6.4 by defining $\mathbb{E}_1 = [I_N, 0]$. The covariance Ω_N of $(\mathbb{Y})_s$ is then given as:

$$\Omega_N = \mathbb{E}_1 \Omega_p \mathbb{E}_1^T$$

Note that Ω_N is the first $N \times N$ block of Ω_p. Also note that we have,

$$\tilde{\mathcal{A}} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} = (\Omega_N)_s = (\mathbb{E}_1 \otimes \mathbb{E}_1)(\Omega_p)_s = (\mathbb{E}_1 \otimes \mathbb{E}_1)\mathcal{A} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix}$$

$$\therefore \tilde{\mathcal{A}} = (\mathbb{E}_1 \otimes \mathbb{E}_1)\mathcal{A}$$

Using the above equations, we simplify the necessary conditions in Equation 6.13 to get:

$$\mathcal{A}^T (\mathbb{E}_1^T \otimes \mathbb{E}_1^T) W^{-1} (\mathbb{E}_1 \otimes \mathbb{E}_1) \mathcal{A} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \mathcal{A}^T (\mathbb{E}_1^T \otimes \mathbb{E}_1^T) W^{-1} (\mathbb{Y}\mathbb{Y}^T)_s = 0 \tag{6.15}$$

Using an analysis similar to the previous section, to get the MLE necessary conditions in Equation 6.9 to be equivalent to the conditions in Equation 6.15 we need W defined to satisfy the following condition:

$$(\mathbb{E}_1^T \otimes \mathbb{E}_1^T) W^{-1} (\mathbb{E}_1 \otimes \mathbb{E}_1) = (I_{pN_d} \otimes \Omega_p^{-2})$$

The weighted ALS conditions in Equation 6.15 then becomes:

$$\mathcal{A}^T (\Omega_p^{-1})_s - \mathcal{A}^T (I_{pN_d} \otimes \Omega_p^{-2}) M (\mathbb{Y}\mathbb{Y}^T)_s = 0$$

Comparing the above Equation with the MLE estimation Equation 6.9 we see that M is some permutation matrix containing 1's and 0 such that it satisfies:

$$M (\mathbb{Y}\mathbb{Y}^T)_s = (Y_p Y_p^T)_s \tag{6.16}$$

Notice that $Y_p Y_p^T$ contains autocovariances for lags up to N_d while $\mathbb{Y}\mathbb{Y}^T$ contains autocovariances for lags only up to the window size N. Thus, from the above equation it is clear that there exists no permutation matrix M that satisfies Equation 6.16. Thus the parameter N cannot be introduced to give an equivalent ALS weight when doing maximum likelihood estimation.

6.2 Conclusions

The application of the maximum likelihood estimation procedure was not tested on actual data sets due to large memory requirements even for problems of small dimensions. We however, expect the performance of the ML formulation to be similar to the

weighted least-squares formulation of the ALS problem given in Section 4.1. Solving the ALS problem with $N = N_d$ and iterating till convergence is a practical method of implementation of the maximum likelihood formulation provided the computation is feasible.

Chapter 7

Industrial Applications of the ALS Technique and Extension to Nonlinear Models [1]

Most physical systems have nonlinear dynamics when modelled from first principles. Even a simple reaction such as $A + B \rightarrow C$ has a nonlinear rate of reaction $r = kC_AC_B$ and the rate constant k has a nonlinear dependence on the temperature. A linear time-invariant approximation to the nonlinear model is often not valid when the dynamics are significantly nonlinear or the plant is operated under batch mode eg. a co-polymerization reactor requiring batch transitions. Techniques that have been developed for linear time-invariant models then have to be extended to nonlinear models and their validity tested.

[1] **Portions of this chapter are to appear in Rajamani et al. (October, 2007)**

In this chapter, we present the nonlinear extension to the Autocovariance Least-Squares (ALS) technique developed in Chapters 3 and 5. A time-varying ALS formulation is applied to real industrial data from a nonlinear blending drum used at *ExxonMobil Chemical Company* [2]. Section 7.2 presents the model and the results to illustrate the value of the nonlinear extension to the ALS technique. Section 7.3 shows the equivalence between a current industrial state estimator implementation called the Implicit Dynamic Feedback (IDF) and integrating input disturbance models. The results from Chapter 5 are used in Section 7.6 to identify appropriate covariances for the disturbance structure in *Shell's* SMOC implementation.

7.1 Noise Covariance Estimation for Nonlinear Models

Consider the following nonlinear state-space model:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + g(x_k, u_k) w_k \\ y_k &= h(x_k) + v_k \end{aligned} \tag{7.1}$$

where, $x \in \mathbb{R}^n$ is the state, $y \in \mathbb{R}^p$ are the measurements, $u_k \in \mathbb{R}^m$ are the inputs w_k, v_k and the noises corrupting the state and the measurements respectively. The noises w_k, v_k are assumed to be from a Gaussian distribution with mean zero and covariances Q_w, R_v. Also, w_k, v_k are assumed to be uncorrelated with each other.

Let the state estimates be obtained starting from an arbitrary initial value using a time-varying stable filter gain sequence L_k. An example of such a filter gain would

[2] *ExxonMobil Chemical Company* is a division of *ExxonMobil Corporation*

be the sequence obtained by implementing the extended Kalman filter (EKF). The only condition on the time-varying gains L_k is that they are stable (see Assumption 7.1). The state estimation can then be described by the following equations:

$$\begin{aligned} \hat{x}_{k+1|k} &= f(\hat{x}_{k|k}, u_k) \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} - L_k(y_k - \hat{y}_{k|k-1}) \\ \hat{y}_{k|k-1} &= h(\hat{x}_{k|k-1}) \end{aligned} \tag{7.2}$$

Here, $\hat{x}_{k|k-1}$ represents the predicted estimate of the state x_k given measurements and inputs up to time t_{k-1} and $\hat{x}_{k|k}$ the filtered estimate given measurements and inputs up to time t_k. Subtracting the predicted state estimates in Equation 7.2 from the plant in Equation 7.1, we get an approximate time-varying linear model for the innovations:

$$\begin{aligned} \varepsilon_{k+1} &\approx \underbrace{(A_k - A_k L_k C_k)}_{\bar{A}_k} \varepsilon_k + \underbrace{\begin{bmatrix} G_k & -A_k L_k \end{bmatrix}}_{\bar{G}_k} \underbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}_{\bar{w}_k} \\ y_k &\approx C_k \varepsilon_k + v_k \end{aligned} \tag{7.3}$$

where, $\varepsilon_k = (x_k - \hat{x}_{k|k-1})$ denotes the state estimate error and y_k the innovations. The noises w_k, v_k driving the innovations sequence are assumed to drawn from time-invariant covariances Q_w, R_v.

The time-varying approximation of the nonlinear model is defined by the following linearization:

$$A_k = \left.\frac{\partial f(x_k, u_k)}{\partial x_k^T}\right|_{(\hat{x}_{k|k-1}, u_k)}, \quad G_k = g(\hat{x}_{k|k-1}, u_k), \quad C_k = \left.\frac{\partial h(x_k)}{\partial x_k^T}\right|_{\hat{x}_{k|k-1}} \tag{7.4}$$

A second possibility for the linear approximation is to evaluate the terms in Equation 7.4 at $\hat{x}_{k|k}$ instead of $\hat{x}_{k|k-1}$ (Anderson and Moore, 1979). The next section presents a technique for estimating the covariances Q_w, R_v using autocovariances of data at different time lags.

7.1.1 Time-varying Autocovariance Least-Squares Technique

The Autocovariance Least-Squares (ALS) covariance estimation technique described in Chapter 3 was developed for linear time-invariant models. When using nonlinear or time-varying models, a key difference is that the estimate error covariance $P_k = E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})]$ is the time-varying solution to the Riccati equation and does not reach a steady state value. No simple equation can then be written for P_k in terms of Q_w, R_v and the system matrices as in the linear time-invariant case. The following Assumption 7.1 however, allows extension of the ALS technique to time-varying and nonlinear systems.

Assumption 7.1. *The time-varying filter gain sequence L_k used in Equation 7.2 are such that when used in the approximate linearization given by Equation 7.3 they produce a sequence of $\bar{A}_k = (A_k - A_k L_k C_k)$ matrices such that the product $(\prod_{k=0}^{j} \bar{A}_i) \approx 0$ as j increases.*

Starting from an arbitrary initial condition ε_0 at t_0, consider the evolution of

Equation 7.3 up to time t_k to obtain the following:

$$\begin{aligned}
\mathcal{Y}_k &= C_k(\prod_{i=0}^{k-1} \bar{A}_i)\varepsilon_0 + C_k(\prod_{i=1}^{k-1} \bar{A}_i\bar{G}_0\bar{w}_0 + \prod_{i=2}^{k-1} \bar{A}_i\bar{G}_1\bar{w}_1 + \cdots + \bar{G}_{k-1}\bar{w}_{k-1}) + v_k \\
\mathcal{Y}_{k+j} &= C_{k+j}(\prod_{i=0}^{k+j-1} \bar{A}_i)\varepsilon_0 + C_{k+j}(\prod_{i=1}^{k+j-1} \bar{A}_i\bar{G}_0\bar{w}_0 + \prod_{i=2}^{k+j-1} \bar{A}_i\bar{G}_1\bar{w}_1 + \cdots + \bar{G}_{k+j-1}\bar{w}_{k+j-1}) + v_{k+j}
\end{aligned} \tag{7.5}$$

The covariance of $\bar{w}_k = \begin{bmatrix} w_k \\ v_k \end{bmatrix}$ is given by $\bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}$.

From Assumption 7.1, the index $k > 0$ is chosen to be large enough such that $\prod_{i=0}^{k-1} \bar{A}_i \approx 0$. The effect of the initial estimate error ε_0 is then negligible in $\mathcal{Y}_{k+j}$ for $j > 0$. The terms involving ε_0 in Equation 7.5 can then be neglected. From Equation 7.5 we get the following expressions for the expectation of the autocovariances at different lags:

$$\begin{aligned}
E[\mathcal{Y}_k\mathcal{Y}_k^T] &= C_k(\prod_{i=1}^{k-1} \bar{A}_i\bar{G}_0 + \cdots + \bar{G}_{k-1})\bar{Q}_w(\bar{G}_0^T\prod_{i=1}^{k-1} \bar{A}_i^T + \cdots + \bar{G}_{k-1}^T)C_k^T + R_v \\
E[\mathcal{Y}_{k+j}\mathcal{Y}_k^T] &= C_{k+j}(\prod_{i=1}^{k+j-1} \bar{A}_i\bar{G}_0 + \cdots + \prod_{i=k}^{k+j-1} \bar{A}_i\bar{G}_{k-1})\bar{Q}_w(\bar{G}_0^T\prod_{i=1}^{k-1} \bar{A}_i^T + \cdots + \bar{G}_{k-1}^T)C_k^T \\
&\quad - C_{k+j}(\prod_{i=k+1}^{k+j-1} \bar{A}_i)A_kL_kR_v
\end{aligned} \tag{7.6}$$

In the above Equation 7.6, we define $\prod_{i=k}^{j} \bar{A}_i = I$ for all $j \leq k$.

Remark 7.1. *In general, instead of starting from the initial condition ε_0 at t_0, we can start from an initial condition ε_m at t_m and calculate the expectations as above, provided the*

indices k and m are such that the following condition holds as in Assumption 7.1:

$$\prod_{i=m}^{m+k-1} \bar{A}_i \approx 0$$

Let the autocovariance matrix $\mathcal{R}_k(N)$ be defined as the expectation of the innovations data at different time lags over a user defined window N (Jenkins and Watts, 1968).

$$\mathcal{R}_k(N) = E\begin{bmatrix} y_k y_k^T \\ \vdots \\ y_{k+N-1} y_k^T \end{bmatrix} \tag{7.7}$$

Using Equations 7.5 and 7.7, we get:

$$\mathcal{R}_k(N) = \underbrace{\begin{bmatrix} I \\ -C_{k+1}A_kL_k \\ \vdots \\ -C_{k+N-1}(\prod_{i=k+1}^{k+N-2}\bar{A}_i)A_kL_k \end{bmatrix}}_{\Psi_1} R_v + \Gamma\Omega_1 \bigoplus_{i=1}^{k} Q_w \Omega_1^T \Gamma_1^T + \Gamma\Omega_2 \bigoplus_{i=1}^{k} R_v \Omega_2^T \Gamma_1^T \tag{7.8}$$

where the matrices are defined as follows and dimensioned appropriately:

$$\Gamma = \begin{bmatrix} C_k(\prod_{i=1}^{k-1}\bar{A}_i) & \cdots & C_k \\ \vdots & \ddots & \vdots \\ C_{k+N-1}(\prod_{i=1}^{k+N-2}\bar{A}_i) & \cdots & C_{k+N-1}(\prod_{i=k}^{k+N-2}\bar{A}_i) \end{bmatrix} \tag{7.9}$$

and

$$\Omega_1 = \begin{bmatrix} G_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & G_{k-1} \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} -A_0L_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -A_{k-1}L_{k-1} \end{bmatrix}, \quad \Gamma_1^T = \begin{bmatrix} (\prod_{i=1}^{k-1}\bar{A}_i^T)C_k^T \\ \vdots \\ C_k^T \end{bmatrix} \tag{7.10}$$

Γ_1 is the first row block of the Γ matrix. In the above equation and in the remainder of this chapter we use standard properties and symbols of Kronecker products (Steeb (1991) and Graham (1981, chap. 2)). $\otimes$ is the standard symbol for the Kronecker product and $\bigoplus$ is the symbol for the Kronecker sum, satisfying the following property:

$$\bigoplus_{i=1}^{k} Q_w = (I_k \otimes Q_w)$$

We use the subscript 's' to denote the column-wise stacking of the elements of a matrix into a vector. Stacking both sides of Equation 7.8 we then get:

$$[\mathcal{R}_k(N)]_s = (\Gamma_1\Omega_1 \otimes \Gamma\Omega_1)\mathcal{I}_{g,k}(Q_w)_s + \left[(\Gamma_1\Omega_2 \otimes \Gamma\Omega_2)\mathcal{I}_{p,k} + I_p \otimes \Psi\right](R_v)_s \tag{7.11}$$

Here, $\mathcal{I}_{p,N}$ is a permutation matrix containing 1's and 0's and satisfying the relation:

$$\left(\bigoplus_{i=1}^{N} R_v\right)_s = \mathcal{I}_{p,N}(R_v)_s$$

If we have an estimate of the autocovariance matrix $\mathcal{R}_k(N)$ denoted by $\hat{\mathcal{R}}_k(N)$ and let $\hat{b}_k = [\hat{\mathcal{R}}_k(N)]_s$, then from Equation 7.11 we can formulate a positive semidefinite constrained least-squares problem in the unknown covariances Q_w, R_v (Chapter 3 and Odelson et al. (2006b)). The optimization to be solved is given by:

$$\Phi_k = \min_{Q_w, R_v} \left\| \mathcal{A}_k \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \hat{b}_k \right\|_W^2 \tag{7.12}$$

$$\text{subject to,} \quad Q_w, R_v \geq 0, \quad Q_w = Q_w^T, R_v = R_v^T$$

where,

$$\mathcal{A}_k = \begin{bmatrix} \mathcal{A}_{k1} & \mathcal{A}_{k2} \end{bmatrix}$$
$$\mathcal{A}_{k1} = (\Gamma_1\Omega_1 \otimes \Gamma\Omega_1)\mathcal{I}_{g,k} \tag{7.13}$$
$$\mathcal{A}_{k2} = \left[(\Gamma_1\Omega_2 \otimes \Gamma\Omega_2)\mathcal{I}_{p,k} + I_p \otimes \Psi\right]$$

We will refer to the optimization in Equation 7.12 as the Autocovariance Least-Squares (ALS) technique. Necessary and sufficient conditions for the uniqueness of the ALS optimization in Equation 7.12 are given in Chapter 3. For the estimates of Q_w, R_v in the ALS optimization to have minimum variance the weighting matrix W in the ALS objective is given by $W = T_k^{-1}$ where T_k is the covariance of $\hat{b}_k$ (Aitken, 1935).

The matrices $\mathcal{A}_k$ and the vector $\hat{b}_k$ in the Equation 7.12 have the time subscript 'k' to emphasize that these quantities are time-varying and based on the time-varying approximation given in Equation 7.3.

Since the only data set available for estimating the time-varying quantity $\mathcal{R}_k(N)$ defined in Equation 7.7 is $\{y_k, \cdots, y_{k+N-1}\}$, the only calculable estimate of $\hat{b}_k$ is given by:

$$\hat{b}_k = [\hat{\mathcal{R}}_k(N)]_s = \begin{bmatrix} y_k y_k^T \\ \vdots \\ y_{k+N-1} y_k^T \end{bmatrix}_s \tag{7.14}$$

At every time instant t_k, we compute the quantities $\mathcal{A}_k$ and $\hat{b}_k$ from Equations 7.13 and 7.14. To simplify the computation, the matrices $\Gamma, \Omega_1, \Omega_2, \Gamma_1$ defined in Equations 7.9 and 7.10 and used in the calculation of $\mathcal{A}_k$, can be computed starting from an

initial condition at time t_m rather than t_0 as given in Remark 7.1. We then use a sliding window strategy to compute the time-varying matrices $\mathcal{A}_k$ and $\hat{b}_k$. Figure 7.1 illustrates the calculation procedure.

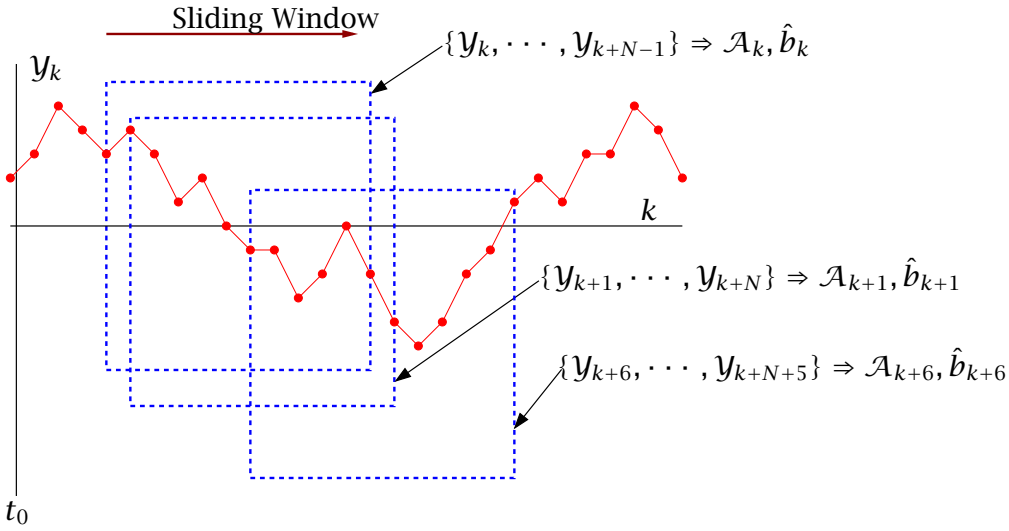


Figure 7.1: Strategy for calculating the time-varying $\mathcal{A}_k$ matrices in Equation 7.13

Using the computed time-varying matrices and the ALS formulation in Equation 7.12, we can then solve the following optimization for a set of data of length N_d to estimate Q_w, R_v:

$$\Phi = \min_{Q_w, R_v} \left\| \begin{bmatrix} \mathcal{A}_k \\ \vdots \\ \mathcal{A}_{N_d-N+1} \end{bmatrix} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \begin{bmatrix} \hat{b}_k \\ \vdots \\ \hat{b}_{N_d-N+1} \end{bmatrix} \right\|_{W_f}^2 \tag{7.15}$$

$$\text{subject to,} \quad Q_w, R_v \geq 0, \quad Q_w = Q_w^T, R_v = R_v^T$$

Since $\hat{b}_k, \hat{b}_{k+1}, \cdots$ are not independent, the weighting matrix W_f is not block diagonal. The formula for W_f is a complicated function of the unknown covariances Q_w, R_v and

an iterative procedure is required for its calculation (see Section 4.1, Chapter 4). We use $W_f = I$ to avoid the expensive and nonlinear calculation.

Remark 7.2. *Notice that if the system is time-invariant we have* $\mathcal{A}_k = \cdots = \mathcal{A}_{N_d-N+1}$ *and we can then recover the time-invariant ALS optimization presented in Section 3.2 in Chapter 3.*

Remark 7.3. *Assumption 7.1 is a simple practical requirement that the time-varying linear approximation of the full nonlinear model has a gain sequence that makes the estimate error asymptotically zero. This requirement is satisfied in most industrial applications that use a linear approximation to design the state estimator.*

7.2 Industrial Blending Drum Example

Figure 7.2 shows a schematic diagram of a blending drum where two components, monomer A and co-monomer B are mixed with a diluent to give the correct proportions for the blends used as the feed for the different grades of polymers made in the reactors. The diagram is a representation of a blending process unit at *ExxonMobil Chemical Company*. The mass fractions of A and B inside the blending drum are represented by X_A and X_B. The blending tank is elliptical at the bottom and the dynamics of the blending process are nonlinear.

The states of the system are the mass fractions X_A, X_B and the level in the drum h. The nonlinearity in the drum is given by the following equations where V is the volume

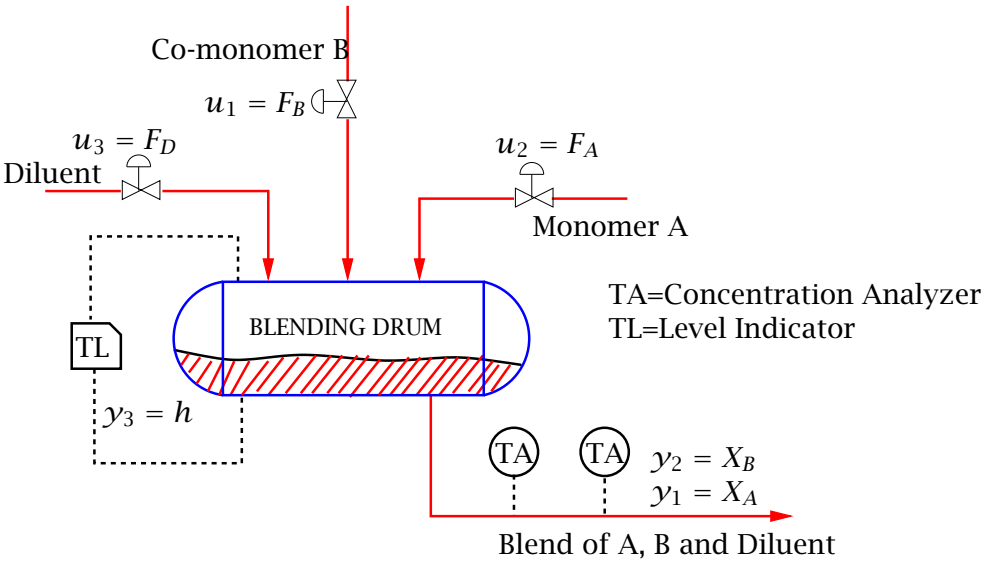


Figure 7.2: Schematic diagram of the blending drum: the level and mass fractions h, X_A, X_B are controlled by flows F_A, F_B, F_D

of the blend in the drum and C_1, C_2, C_3, C_4 are constants:

$$V = C_1 h^3 + C_2 h^2 + C_3 h + C_4$$

$$\frac{dV}{dh} = 3C_1 h^2 + 2C_2 h + C_3$$

A diluent with mass flow rate F_D is added to the blending tank to maintain the required monomer to co-monomer ratio. If the mass flow rates of A and B into the drum are given by F_A and F_B respectively, we can write simple mass balances for the dynamics of the states X_A, X_B, h.

$$\begin{aligned}
\frac{dX_A}{dt} &= \frac{F_A - X_A(F_A + F_B + F_D)}{\rho(C_1 h^3 + C_2 h^2 + C_3 h + C_4)} \\
\frac{dX_B}{dt} &= \frac{F_B + X_{BD} F_D - X_B(F_A + F_B + F_D)}{\rho(C_1 h^3 + C_2 h^2 + C_3 h + C_4)} \\
\frac{dh}{dt} &= \frac{(F_A + F_B + F_D) - F_{out}}{\rho(C_1 h^2 + C_2 h + C_3)}
\end{aligned} \tag{7.16}$$

where, X_{BD} is the mass fraction of B in the diluent feed, F_{out} is the measured out flow rate of the blend and ρ is the recorded density of the blend.

The controlled variables are the level of inventory in the drum h, the ratio of the mass fractions of A to B, $\frac{X_A}{X_B}$ in the outlet stream and the fraction X_A of A leaving the drum. Measurements are made for the states X_A, X_B, h at intervals of one minute.

Five sets of industrial operating data were provided by *ExxonMobil Chemical Company* with the characteristics described below:

- **Data Set I:** Steady state operating data
- **Data Set II:** Set point change in the liquid level h
- **Data Set III:** Step change in the in flow rate F_B
- **Data Set IV:** Set point change in the liquid level and the ratio of the monomer to co-monomer mass fractions
- **Data Set V:** Same as Data Set IV

The state-space model in discrete time for the above equations can then be represented as:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + Gw_k \\ y_k &= x_k + v_k \end{aligned} \tag{7.17}$$

in which,

$$x_k = \begin{bmatrix} X_A \\ X_B \\ h \end{bmatrix}, \quad u_k = \begin{bmatrix} F_A \\ F_B \\ F_D \end{bmatrix}$$

The G matrix is chosen as the constant $G = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ in the model, which implies the state noise in h is more significant than the noise in the states X_A, X_B. The state x_{k+1} is calculated by integrating the nonlinear model given in Equation 7.16 from x_k over the discretization time of one minute while keeping the input u_k constant. The states and the measurements are corrupted by noises w_k, v_k with unknown covariances Q_w, R_v respectively.

7.2.1 Simulation Results

To illustrate the value of the ALS technique applied to a nonlinear model, the data were generated by drawing noises w_k and v_k from known covariances Q_w, R_v and simulating the nonlinear model from Equations 7.16 as given in Equation 7.17. Since a steady state simulation is equivalent to using a linear time-invariant approximation to the nonlinear model, the data are simulated by making multiple set point changes to the level in the drum and implementing a simple PI controller to maintain the level. The set point changes introduced are such that a simple linear time-invariant approximation of the full nonlinear model is not valid.

Figure 7.3 shows a plot of the eigenvalues of the time-varying $(A_k - A_k L_k C_k)$

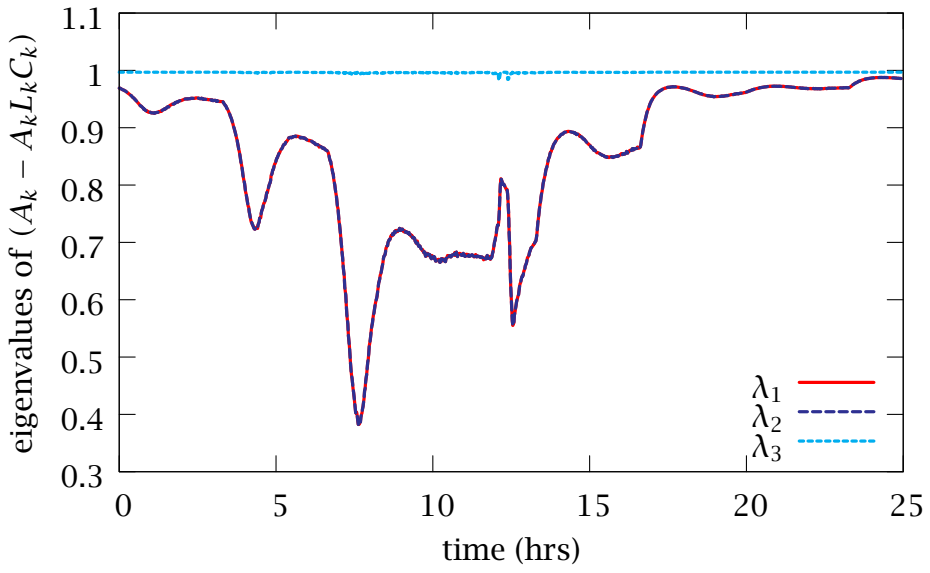


Figure 7.3: Eigenvalues of $(A_k - A_k L_k C_k)$ plotted against time for the simulated data

matrices from the simulation as a function of time. From the plot in Figure 7.3 it is clear that a linear time-invariant approximation of the nonlinear model is not applicable.

The covariances in the simulated data are chosen as:

$$Q_w = 2 \times 10^{-5}, \qquad R_v = \begin{bmatrix} 2 \times 10^{-9} & 0 & 0 \\ 0 & 3.2 \times 10^{-7} & 0 \\ 0 & 0 & 3 \times 10^{-3} \end{bmatrix}$$

The ALS technique with the time-varying strategy described in Section 7.1 is then applied to the simulated data to recover the noise covariances.

Figure 7.4 shows a scatter plot of the covariances estimated using the ALS technique repeated 200 times with $N_d = 200$ and $N = 15$. The mean of the estimates is

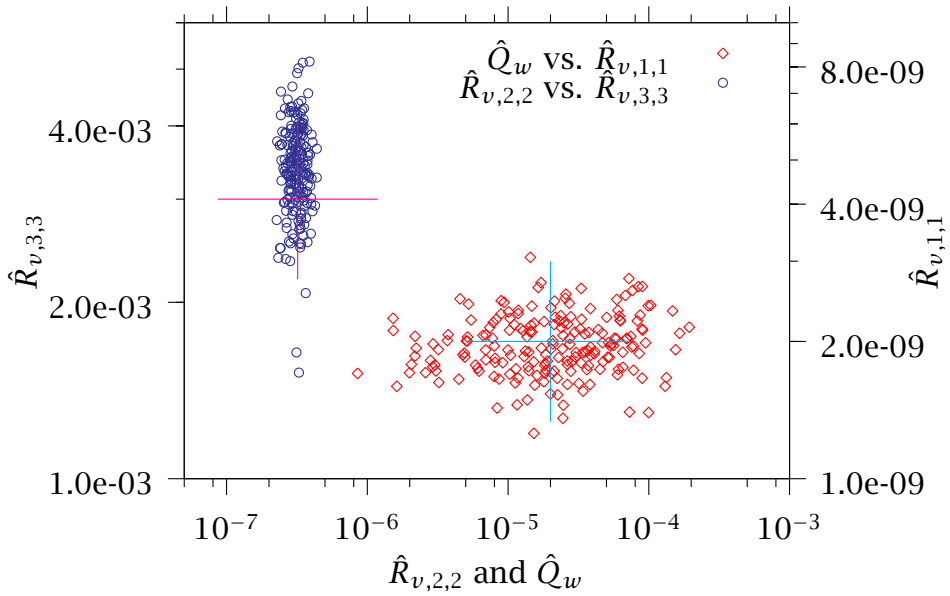


Figure 7.4: Scatter in the covariance estimates using the ALS technique on simulated data. The + sign shows the actual value of the covariances

calculated as:

$$\hat{Q}_w = 3.23 \times 10^{-5}, \qquad \hat{R}_v = \begin{bmatrix} 1.98 \times 10^{-9} & -2.71 \times 10^{-11} & 6.01 \times 10^{-9} \\ -2.71 \times 10^{-11} & 3.17 \times 10^{-7} & -1.12 \times 10^{-7} \\ 6.01 \times 10^{-9} & -1.12 \times 10^{-7} & 3.42 \times 10^{-3} \end{bmatrix}$$

The variance in the estimates seen in the Figure 7.4 decreases as N_d is increased. The covariance estimates using the ALS technique are thus close to the actual covariances used in the simulation.

7.2.2 Using Real Industrial Data

The ALS technique is next applied to five operating data sets provided by *ExxonMobil Chemical Company*. To apply the ALS technique, the innovations sequence defined as

the difference between the measured outputs and the simulated model outputs should be zero mean. To remove the modelling errors and to ensure zero mean in the innovations, integrating disturbances are added to the inputs and the outputs in the model (Tenny et al., 2004; Pannocchia and Rawlings, 2002). A state estimator is then used to estimate the integrating disturbance vector, which then compensates for the plant/model mismatch. To ensure detectability of the integrating disturbance model the number of disturbances is chosen as $n_d = 3$ (Muske and Badgwell, 2002; Pannocchia and Rawlings, 2002) (Also see Remark 7.4).

Remark 7.4. *Note that the level in the tank is an integrator and the model of the plant is not perfect. Hence, when the inputs provided in the data sets is used in the model simulation, the simulated $\hat{y}_{k,3}$ and the measured h from the data can never match. This issue is resolved by adding an integrating input disturbance to the state h. Also note that since the dynamics for h are integrating in the open-loop, an output disturbance model cannot be used to remove this mismatch (Qin and Badgwell, 2003; Dougherty and Cooper, 2003).*

The model in Equation 7.17 augmented with the disturbance model is given by:

$$x_{k+1} = f(x_k, u_k) + B_d d_k + G w_k \tag{7.18}$$

$$d_{k+1} = d_k + \xi_k \tag{7.19}$$

$$y_k = x_k + C_d d_k + v_k \tag{7.20}$$

where, d_k is the integrating disturbance and $\xi \sim N(0, Q_\xi)$ is the noise model for the

disturbance. We make the following choice for B_d and C_d:

$$B_d = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad C_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The model is simulated with the inputs and measurements given in the data sets and the extended Kalman filter (EKF) used as the initial state estimator. Since the covariances affecting the operating data are unknown, the following initial guesses for the covariances are used to calculate the initial estimator gain sequence in the EKF:

$$Q_w = 1 \times 10^{-14}, \quad Q_\xi = 10^{-9} \times I_3, \quad R_v = 2 \times \text{diag}\,(10^{-7}, 10^{-5}, 10^{-3}) \tag{7.21}$$

These choices are made by assigning the variance in the operating data measurements to the corresponding diagonal elements of R_v and choosing small values for Q_w, Q_ξ. This choice follows the rough industrial guidelines of assigning the noise covariances to the output when calculating state estimator gains.

The initial estimator gain sequence L_k obtained by implementing the EKF satisfies Assumption 7.1 and hence the ALS technique can be applied to the operating data. With the above model and the industrial data, the covariances Q_w, Q_ξ and R_v are estimated. The diagonal elements of the estimated covariances Q_w, Q_ξ, R_v using the ALS technique for the five data sets are shown in Figures 7.5 and 7.6. The figures show that the estimated covariances have consistent values across all the five industrial data sets provided. These estimated covariances are then used to calculate new filter gains for the EKF. A snapshot of the measurements compared against the model estimates with

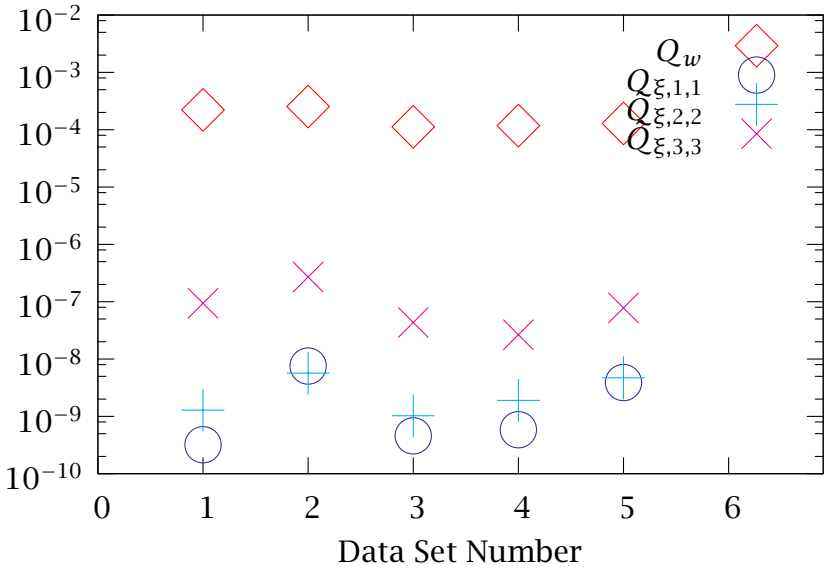


Figure 7.5: Comparison of the diagonal elements of Q_w and Q_ξ for the 5 industrial data sets

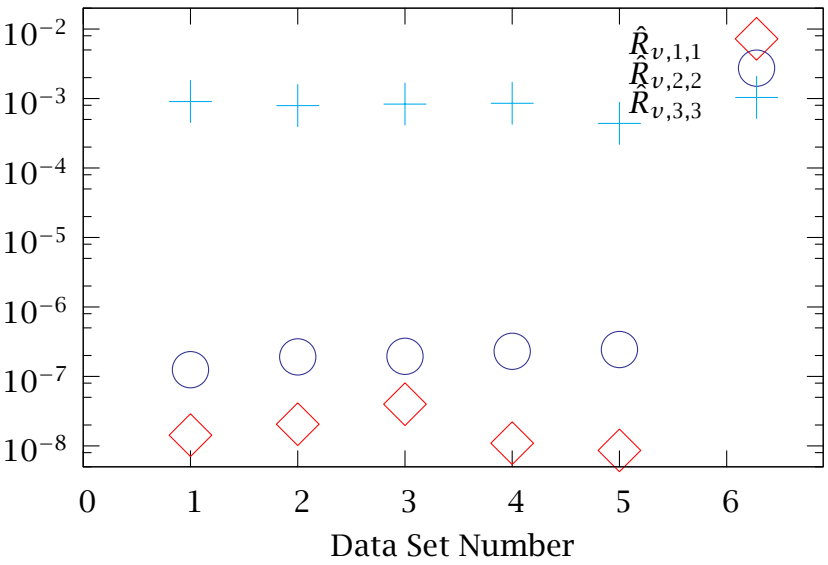


Figure 7.6: Comparison of the diagonal elements of R_v for the 5 industrial data sets

the initial and the ALS covariances for Data Set I is shown in Figures 7.7, 7.8 and 7.9. The y-axis on the figures have been arbitrarily scaled to disguise the original industrial operating data. The figures also show the estimated disturbance vector $\hat{d}_k$ with the initial and the ALS covariances.

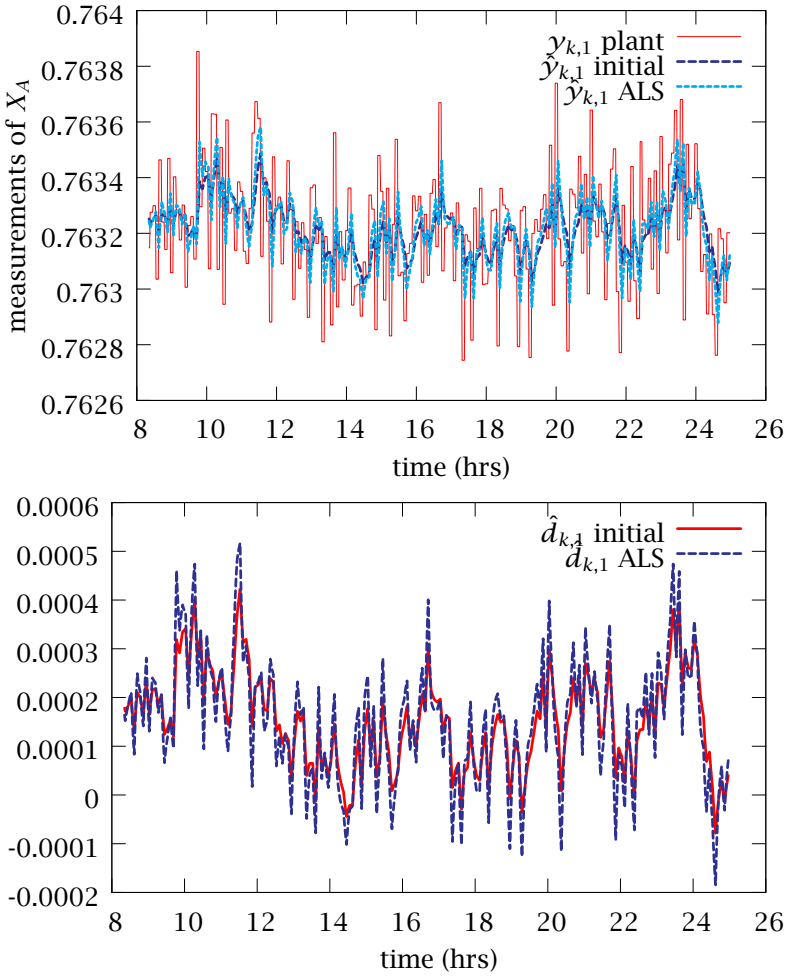


Figure 7.7: A snapshot of data comparing the model estimates using the initial and the ALS covariances with the operating data in Data Set I (the y-axis is arbitrarily scaled to disguise the original data)

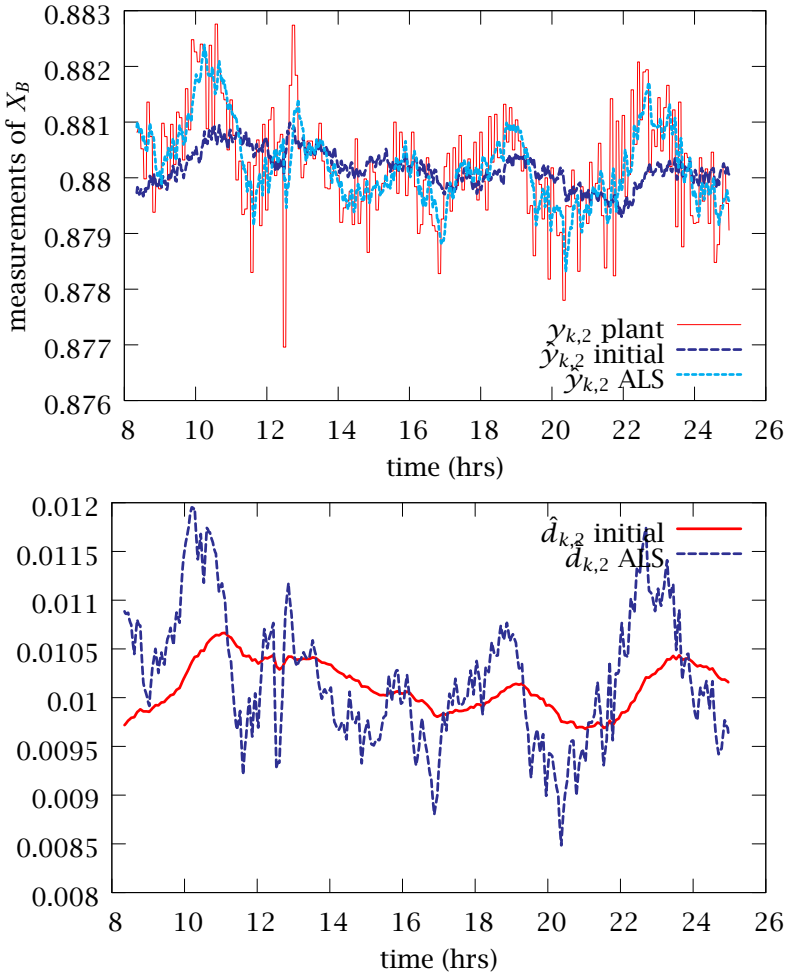


Figure 7.8: A snapshot of data comparing the model estimates using the initial and the ALS covariances with the operating data in Data Set I (the y-axis is arbitrarily scaled to disguise the original data)

Remark 7.5. *If the operating data is collected close to the steady state, the time-varying linear approximation of the nonlinear model given in Equation 7.3 can be further simplified to a linear time-invariant model since the linearized terms in Equation 7.4 are evaluated close the steady state and have approximately the same values. When there*

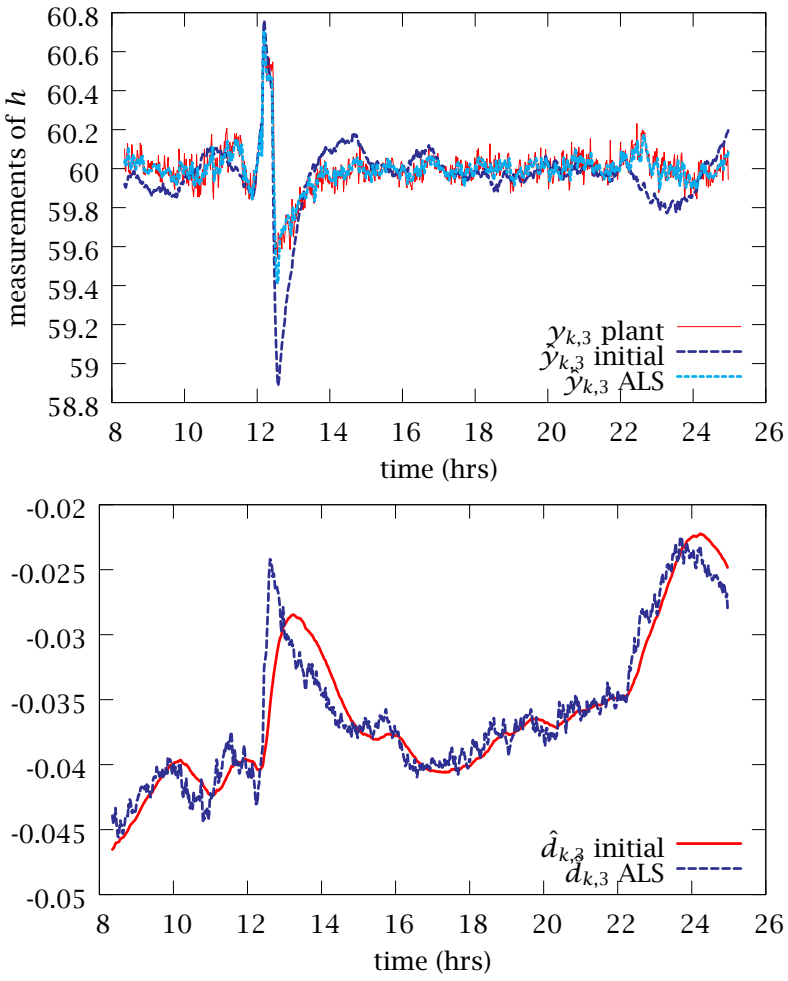


Figure 7.9: A snapshot of data comparing the model estimates using the initial and the ALS covariances with the operating data in Data Set I (the y-axis is arbitrarily scaled to disguise the original data)

are multiple set point changes in the operating data, the transience can be ignored and the linearization at the new set point can be approximated as a new time-invariant linear model. This time-invariant approximation is useful to compare the state estimator specified with the ALS technique against other choices when the actual states are unknown.

A check for the optimality of the state estimator is to test the correlations of the innovations $y_k = (\mathcal{Y}_k - h(\hat{x}_{k|k-1}))$ for whiteness and zero mean (Gelb, 1974, p. 319). For the linear time-invariant model, when the optimal state estimator gains are implemented, the autocovariance matrix in Equation 7.7 is given as:

$$\mathcal{R}_k(N) = \begin{bmatrix} CPC^T + R_v \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where, P is the time-invariant solution to the Ricatti equation with the actual covariances Q_w, R_v. Note that the lagged autocovariances are zero.

We use a similar approach to test the state estimator specified with the covariances Q_w, Q_ξ, R_v estimated using the ALS technique. Following Remark 7.5 and the above analysis of the autocovariances for the time-invariant model, we only consider innovations of the steady state operating data and neglect the transience between the set point changes. If the values of the estimated covariances are close to the actual covariances in the operating data, we expect the innovations approximated from a linear time-invariant model to be uncorrelated in time and the lagged autocovariances of the data to be close to zero.

Figure 7.10 shows a plot comparing the autocovariances estimated from steady state data at different time lags using filter gains from the ALS technique and the initial covariances in Equation 7.21. As seen in the figure, the values of the lagged autoco-

variances are close to zero after a time lag of 4 and remain zero with increasing time lags when the ALS covariance estimates are used. On the other hand, the lagged autocovariances remain nonzero when the state estimator is specified with the rough initial covariances. Clearly the identified covariances with the integrating disturbance model have improved the accuracy of the state estimates. We can also see there remains some undermodelling of the plant disturbance given the nonzero autocovariance at small lags.

An estimate of $E[y_k y_k^T] = CPC^T + R_v$ for Data Set I, calculated using a time-invariant linear approximation by averaging the time-varying matrices and covariances $\hat{Q}_w, \hat{Q}_\xi, \hat{R}_v$ specified using the ALS technique is given by:

$$C_m \hat{P} C_m^T \approx \text{diag}\,(2.3 \times 10^{-9}, 1.3 \times 10^{-8}, 6.0 \times 10^{-4})$$

$$\hat{R}_v \approx \text{diag}\,(1.43 \times 10^{-8}, 1.25 \times 10^{-7}, 9.05 \times 10^{-4})$$

$$\text{giving,} \qquad C_m \hat{P} C_m^T + \hat{R}_v \approx \text{diag}\,(1.6 \times 10^{-8}, 1.4 \times 10^{-7}, 1.5 \times 10^{-3})$$

where, C_m is the average of the time-varying linearization C_k for Data Set I and $\hat{P}$ is calculated as the solution of the Riccati equation from the average time-varying system. These values are in good agreement with the zero lag autocovariances shown in Figure 7.10.

The plots for the other data sets are similar and presented in Appendix 7.7.1.

7.2.3 Conclusions

The Autocovariance Least-Squares (ALS) procedure was extended to nonlinear models to estimate the noise covariances from operating data. A semidefinite constrained opti-

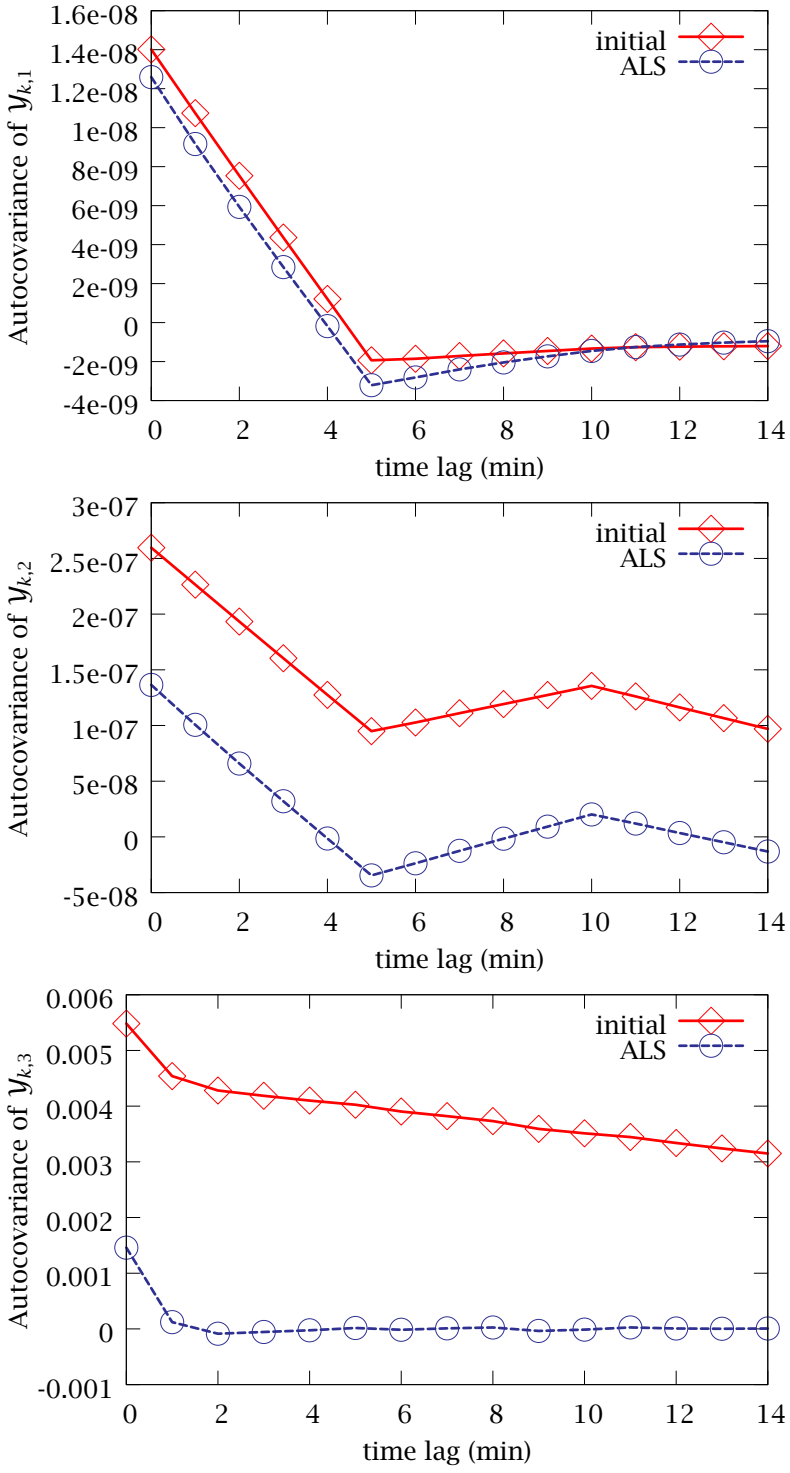


Figure 7.10: Data Set I: Diagonal elements of the innovations autocovariance matrix at different time lags

mization with a least-squares objective was solved to give the covariance estimates. The ALS technique was then shown to give good covariance estimates from data simulated with known noise statistics. The practical benefits of the technique was illustrated by an application to blending drum operating data sets provided by *ExxonMobil Chemical Company*. Consistent values for the covariances were estimated in all the industrial data sets provided by *ExxonMobil Chemical Company*. The state estimator specified with covariances calculated using ALS technique produced white innovations implying optimality in comparison to an initial choice for the covariances. The ALS technique was also used to estimate the covariances for integrating output disturbance models to remove the plant/model mismatch. The improvement in the estimator performance implies cost benefits in the implementation of advanced control schemes (see for eg. Odelson et al. (2006a) for numerical values) when the ALS technique is used to estimate the noise covariances.

7.3 Connections between IDF and Using Integrating Input Disturbance Models

Consider the plant to evolve according to the following continuous time nonlinear state-space model:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) + g(x(t), u(t))w \\ y(t) &= h(x(t)) + v \end{aligned} \tag{7.22}$$

where $x \in R^n$ is the state, u the input and $y \in R^p$ the measurements. w and v are the process and the measurement noises.

Given the history of measurements $Y_k = \{y_0, y_1, \cdots, y_k\}$ at times $\{t_0, t_1, \cdots, t_k\}$ and a model for the plant, we want to estimate the state of the system at time t_k.

7.4 Implicit Dynamic Feedback (IDF)

Implicit Dynamic Feedback (IDF) is one of the current state estimation techniques used in the chemical industry. IDF is also the state estimator used for the ethylene co-polymerization reactor at *ExxonMobil Chemical Company*.

Let measurements be available at the discrete times $\{t_0, t_1, \cdots\}$. When the measurement at t_{k+1} becomes available, IDF estimates the disturbance variables as the following integral of the difference between the actual plant output and the model output (Hedengren et al., 2007):

$$\hat{d}(t \mid Y_{k+1}) = K(y_{k+1} - \hat{y}(t \mid Y_{k+1})) + \frac{K}{\tau_I} \int_0^t (y_{k+1} - \hat{y}(T \mid Y_{k+1}))dT \tag{7.23}$$

where, $\hat{d}(t \mid Y_{k+1})$ is the estimate of the disturbance given the history of the measurements Y_k. The estimate $\hat{d}(t \mid Y_{k+1})$ is calculated at every time instant between $t_k < t < t_{k+1}$ after the measurement y_{k+1} becomes available. The predicted output $\hat{y}(t \mid Y_{k+1})$ is obtained using the measurement model $y = h(x)$ and by solving the state equation simultaneously. K and τ_I are scalars chosen for each measurement in

the y vector. For example if the measurement vector is of size 3, then K is the matrix:

$$\begin{bmatrix} K_1 & 0 & 0 \\ 0 & K_2 & 0 \\ 0 & 0 & K_3 \end{bmatrix}$$

Each element from the estimated $\hat{d}$ vector is paired with a state of the system. Examples of the pairings are: the measurement *ethylene feed* is paired with *ethylene mol %* which is state variable and measured state *catalyst activity* is paired with the state variable *production rate*.

The evolution of the IDF model over the time interval $t_k < t \leq t_{k+1}$ is:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u_k) + B_d d(t) \\ y(t) &= h(x(t)), \qquad t_k < t \leq t_{k+1} \\ x(t_k) &= \hat{x}_{k|k} \end{aligned} \tag{7.24}$$

in which, B_d is the matrix containing the chosen pairings between the disturbance vector d and the model states x. The input u_k is held constant over the time interval. The notation $x_{k|k}$ is used as a shorthand for $x(t_k \mid Y_k)$.

Thus, to get the IDF state estimate, Equation 7.23 is solved simultaneously with Equation 7.24:

$$\hat{x}(t \mid Y_{k+1}) = \hat{x}_{k|k} + \int_{t_k}^{t} [f(\hat{x}(T), u_k) + B_d \hat{d}(T)]dT, \quad t_k < t \leq t_{k+1} \tag{7.25}$$

Differentiating Equation 7.23 w.r.t time, we get:

$$\frac{\delta \hat{d}(t \mid Y_{k+1})}{\delta t} = -K \left.\frac{dh}{dx}\right|_{\hat{x}(t)} [f(\hat{x}(t), u_k) + B_d \hat{d}(t)] + \frac{K}{\tau_I}(y_{k+1} - h(\hat{x}(t))) \tag{7.26}$$

Thus, to get to IDF state estimates in the time interval $t_k < t \leq t_{k+1}$, after the measurement y_{k+1} is available, the following set of differential equations have to solved simultaneously:

$$\begin{aligned}
\frac{\delta \hat{x}(t \mid Y_{k+1})}{\delta t} &= f(\hat{x}(t), u_k) + B_d \hat{d}(t) \\
\frac{\delta \hat{d}(t \mid Y_{k+1})}{\delta t} &= -K \left.\frac{dh}{dx}\right|_{\hat{x}(t)} [f(\hat{x}(t), u_k) + B_d \hat{d}(t)] + \frac{K}{\tau_I}(y_{k+1} - h(\hat{x}(t)))
\end{aligned} \tag{7.27}$$

subject to initial conditions,

$$\begin{aligned}
\hat{x}(t_k) &= \hat{x}_{k|k} \\
\hat{d}(t_k) &= \hat{d}_{k|k}
\end{aligned}$$

7.5 Equivalence between integrating disturbance models and IDF for discrete-time systems

For purposes of illustrating the equivalence between integrating disturbances and IDF, we consider the following linear state-space model:

$$\begin{aligned}
x_{k+1} &= Ax_k + Bu_k + Gw_k \\
y_k &= Cx_k + v_k
\end{aligned} \tag{7.28}$$

The more general nonlinear, continuous time case is considered briefly in Section 7.5.3.

7.5.1 Input Disturbance Models

To prevent offset in the outputs, model predictive control (MPC) includes integrating disturbance models to account for plant/model mismatch. The integrating disturbance is either added to the input or the output (Pannocchia and Rawlings, 2002). In this report we show the equivalence of IDF and the integrating input disturbance model.

The dynamics of the integrating white noise is included in the state-space model by augmenting the state vector with the disturbance vector m. The state-space model with the integrating disturbance added to the input is then given by:

$$\begin{aligned}\begin{bmatrix} x \\ m \end{bmatrix}_{k+1} &= \begin{bmatrix} A & \bar{B}_d \\ 0 & I \end{bmatrix} \begin{bmatrix} x \\ m \end{bmatrix}_k + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k + \begin{bmatrix} G & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} w \\ \xi \end{bmatrix}_k \\ y_k &= \begin{bmatrix} C & 0 \end{bmatrix} \begin{bmatrix} x \\ m \end{bmatrix}_k + v_k \end{aligned} \tag{7.29}$$

$\bar{B}_d$ models the integrating disturbance m entering the state. The optimal state estimator for this linear augmented model with no constraints is the Kalman filter. If $w_k \sim N(0, Q_w), v_k \sim N(0, R_v)$ and $\xi_k \sim N(0, Q_\xi)$ and the covariances Q_w, R_v, Q_ξ are known parameters, then the Kalman filter gains L_x, L_d for the state and the disturbance respectively can be easily calculated (Anderson and Moore, 1979).

The prediction equation is then:

$$\begin{bmatrix} \hat{x} \\ \hat{m} \end{bmatrix}_{k+1|k} = \begin{bmatrix} A & \bar{B}_d \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{m} \end{bmatrix}_{k|k} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k \tag{7.30}$$

When the measurement y_{k+1} becomes available, the states are updated using the Kalman filter gains (let $\varepsilon_k = (y_k - \hat{y}_{k|k-1}) = (y_k - C\hat{x}_{k|k-1})$):

$$\begin{bmatrix} \hat{x} \\ \hat{m} \end{bmatrix}_{k+1|k+1} = \begin{bmatrix} \hat{x} \\ \hat{m} \end{bmatrix}_{k+1|k} + \begin{bmatrix} L_x \\ L_d \end{bmatrix} \varepsilon_{k+1} \tag{7.31}$$

Writing the Kalman prediction and update equations in one step, we get:

$$\begin{bmatrix} \hat{x} \\ \hat{m} \end{bmatrix}_{k+1|k+1} = \begin{bmatrix} A & \bar{B}_d \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{m} \end{bmatrix}_{k|k} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k + \begin{bmatrix} AL_x + \bar{B}_d L_d \\ L_d \end{bmatrix} \varepsilon_{k+1} \tag{7.32}$$

7.5.2 IDF for discrete-time models

The continuous time IDF described in section 7.4 is implemented for the discrete-time linear model shown in Equation 7.28. If the discretization time is Δt, then an approximation of Equation 7.23 is ($\varepsilon_k = (y_k - \hat{y}_{k|k-1})$):

$$\hat{d}_{k+1|k+1} = K\varepsilon_{k+1} + \alpha \sum_{j=0}^{k+1} \varepsilon_j \tag{7.33}$$

where,

$$\alpha = \frac{K\Delta t}{\tau_I}$$

Equation 7.33 can be written as:

$$\hat{d}_{k+1|k+1} = (K + \alpha)\varepsilon_{k+1} + \alpha \sum_{j=0}^{k} \varepsilon_j$$

Define:

$$\hat{n}_{k|k} = \alpha \sum_{j=0}^{k} \varepsilon_j \tag{7.34}$$

We get,

$$\begin{aligned} \hat{d}_{k+1|k+1} &= (K+\alpha)\varepsilon_{k+1} + \hat{n}_{k|k} \\ \hat{n}_{k|k} &= \hat{n}_{k-1|k-1} + \alpha\varepsilon_k \end{aligned} \tag{7.35}$$

The IDF state estimate for the discrete linear model similar to Equation 7.27 is given by:

$$\hat{x}_{k+1|k+1} = A\hat{x}_{k|k} + Bu_k + B_d\hat{d}_{k+1|k+1}$$

Using Equation 7.35,

$$\hat{x}_{k+1|k+1} = A\hat{x}_{k|k} + Bu_k + B_d(K+\alpha)\varepsilon_{k+1} + B_d\hat{n}_{k|k} \tag{7.36}$$

Thus, from Equations 7.35 and 7.36, we get:

$$\begin{bmatrix} \hat{x} \\ \hat{n} \end{bmatrix}_{k+1|k+1} = \begin{bmatrix} A & B_d \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{n} \end{bmatrix}_{k|k} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k + \begin{bmatrix} B_dK + B_d\alpha \\ \alpha \end{bmatrix} \varepsilon_{k+1} \tag{7.37}$$

Comparing the Kalman filter in Equation 7.32 and the IDF in Equation 7.37, we see that n_k defined in Equation 7.34 takes the role of the integrating white noise disturbance d_k added to the inputs and B_d is the equivalent of $\bar{B}_d$ in the input disturbance model.

Also, we the have the following relationships:

$$\begin{aligned} \alpha &= L_d \\ B_dK + B_d\alpha &= AL_x + \bar{B}_dL_d \end{aligned} \tag{7.38}$$

Thus, if we estimate the Kalman filter gain from data using the ALS techniques given in Chapter 5 and impose a particular structure on the filter gain estimate, we can calculate the equivalent IDF estimator gain.

7.5.3 IDF on Continuous time systems

When the model is nonlinear and in continuous time, an approach similar to Section 7.5 can be taken for implementing the IDF in the form of a Kalman filter.

Consider the state propagation model to be nonlinear and in continuous time. The measurements are obtained at discrete times. The model then has the following form:

$$\text{System Model:} \quad \dot{x}(t) = f(x(t), t) + g(x(t))w$$

$$\text{Measurement Model:} \quad y_k = h_k(x(t_k)) + v_k$$

The state estimation in this case is achieved by implementing a continuous-discrete linearized Kalman filter (Gelb, 1974, p.189). An equivalence between IDF and the continuous-discrete Kalman filter can again been shown using a derivation similar to the previous section.

7.5.4 Conclusions

It was shown that IDF is a special case of the Kalman filter applied to integrating input disturbance models in MPC. This equivalence motivates using the ALS technique to specify the covariances for the IDF implementation.

7.6 Using ALS with Shell Multivariable Optimization Control (SMOC)

The implementation of *Shell*'s SMOC strategy includes an integrating disturbance model with a structure described in Section 7.6.1 to allow more flexibility in removing offset in the presence of unmodelled disturbances and plant/model mismatch. In this section we show that the special disturbance model structure in SMOC is a particular choice for the structures of B_d and C_d. Also as shown in Chapter 5, the ALS technique can be used to estimate the covariances that specify the state estimator for the SMOC disturbance structure and achieve a closed-loop performance equivalent to other detectable choices for B_d and C_d, also with an ALS specified estimator.

7.6.1 Model Description

Figure 7.11 is a schematic diagram showing the structure of the disturbance models implemented in SMOC. Referring to the model shown in Figure 7.11, the state-space models can be written as the following equations: For the first block:

$$\begin{aligned} x_1(k+1) &= A_1 x_1(k) + B_1 u_1(k) \\ y_1(k) &= C_1 x_1(k) + d_1(k) + D_1 u_1(k) \end{aligned} \tag{7.39}$$

For the second block:

$$\begin{aligned} x_2(k+1) &= A_2 x_2(k) + B_2 u_2(k) \\ x_4(k) &= C_2 x_2(k) + D_2 u_2(k) \end{aligned} \tag{7.40}$$

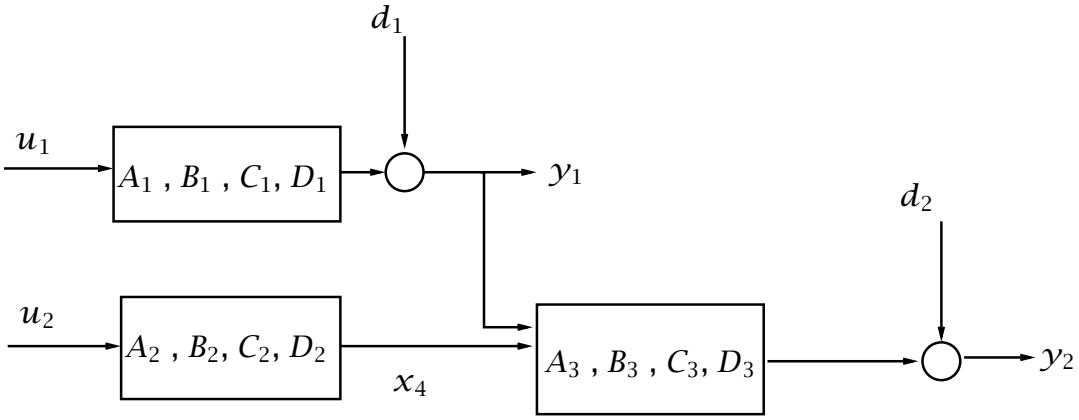


Figure 7.11: Structured SMOC Model

For the third block:

$$
\begin{aligned}
x_3(k+1) &= A_3x_3(k) + B_3(y_1(k) + x_4(k)) \\
x_3(k+1) &= A_3x_3(k) + B_3C_1x_1(k) + B_3d_1(k) + B_3C_2x_2(k) \\
&\quad + B_3D_1u_1(k) + B_3D_2u_2(k) \\
y_2(k) &= C_3x_3(k) + d_2(k) + D_3(y_1(k) + x_4(k)) \\
y_2(k) &= C_3x_3(k) + d_2(k) + D_3C_1x_1(k) + D_3d_1(k) \\
&\quad + D_3D_1u_1(k) + D_3C_2x_2(k) + D_3D_2u_2(k)
\end{aligned}
\tag{7.41}
$$

Putting Equations 7.39- 7.41 together, we get the following:

$$\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \hline d_1 \\ d_2 \end{array}\right]_{k+1} = \left[\begin{array}{ccc|cc} A_1 & 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 & 0 \\ B_3C_1 & B_3C_2 & A_3 & B_3 & 0 \\ \hline 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \hline d_1 \\ d_2 \end{array}\right]_k + \left[\begin{array}{cc} B_1 & 0 \\ 0 & B_2 \\ B_3D_1 & B_3D_2 \\ \hline 0 & 0 \\ 0 & 0 \end{array}\right] \left[\begin{array}{c} u_1 \\ u_2 \end{array}\right]_k$$

$$\left[\begin{array}{c} y_1 \\ y_2 \end{array}\right]_k = \left[\begin{array}{ccc|cc} C_1 & 0 & 0 & I & 0 \\ D_3C_1 & D_3C_2 & C_3 & D_3 & I \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \hline d_1 \\ d_2 \end{array}\right]_k + \left[\begin{array}{cc} D_1 & 0 \\ D_3D_1 & D_3D_2 \end{array}\right] \left[\begin{array}{c} u_1 \\ u_2 \end{array}\right]_k \tag{7.42}$$

Thus, the SMOC disturbance models have the following particular choices for B_d and C_d:

$$B_d = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ B_3 & 0 \end{array}\right], \qquad C_d = \left[\begin{array}{cc} I & 0 \\ D_3 & I \end{array}\right]$$

7.6.2 Examples

To show the equivalence between the SMOC model and other choices for the integrating disturbance model in closed-loop, we choose the following example. Let the transfer functions for the model be given as:

$$\text{First block :} \quad y_1 = \frac{u_1}{2s+1}$$
$$\text{Second block :} x_4 = \frac{u_2}{12s^2+7s+1}$$
$$\text{Third block :} \quad y_2 = \frac{2(y_1+x_4)}{3s+2}$$

A minimal realization of these transfer functions give the following state-space matrices:

$$A_1 = 0.607, \quad A_2 = \begin{bmatrix} 0.966 & 0.747 \\ -0.0623 & 0.530 \end{bmatrix}, \quad A_3 = 0.513$$

$$B_1 = 0.787, \quad B_2 = \begin{bmatrix} 0.413 \\ 0.747 \end{bmatrix}, \quad B_3 = 0.730$$

$$C_1 = 0.5, \quad C_2 = \begin{bmatrix} 0.0833 & 0 \end{bmatrix}, \quad C_3 = 0.667$$

Using the development in the previous section, we get the following for the SMOC disturbance model:

$$A = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ B_3C_1 & B_3C_2 & A_3 \end{bmatrix} = \begin{bmatrix} 0.607 & 0 & 0 & 0 \\ 0 & 0.966 & 0.747 & 0 \\ 0 & -0.0623 & 0.530 & 0 \\ 0.365 & 0.061 & 0 & 0.513 \end{bmatrix}$$

$$B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \\ B_3D_1 & B_3D_2 \end{bmatrix} = \begin{bmatrix} 0.787 & 0 \\ 0 & 0.413 \\ 0 & 0.747 \\ 0 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} C_1 & 0 & 0 \\ D_3C_1 & D_3C_2 & C_3 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.667 \end{bmatrix}$$

$$D = \begin{bmatrix} D_1 & 0 \\ D_3D_1 & D_3D_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The disturbance model added to the SMOC disturbance model is then given as described in the previous section:

$$B_d = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ B_3 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0.730 & 0 \end{bmatrix} \quad C_d = \begin{bmatrix} I & 0 \\ D_3 & I \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Let the alternate disturbance model be an integrating output disturbance model:

$$B_d = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad C_d = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In the plant, let deterministic disturbances enter both the inputs are different intervals as shown in Figure 7.12. The state and sensor noise covariances in the plant

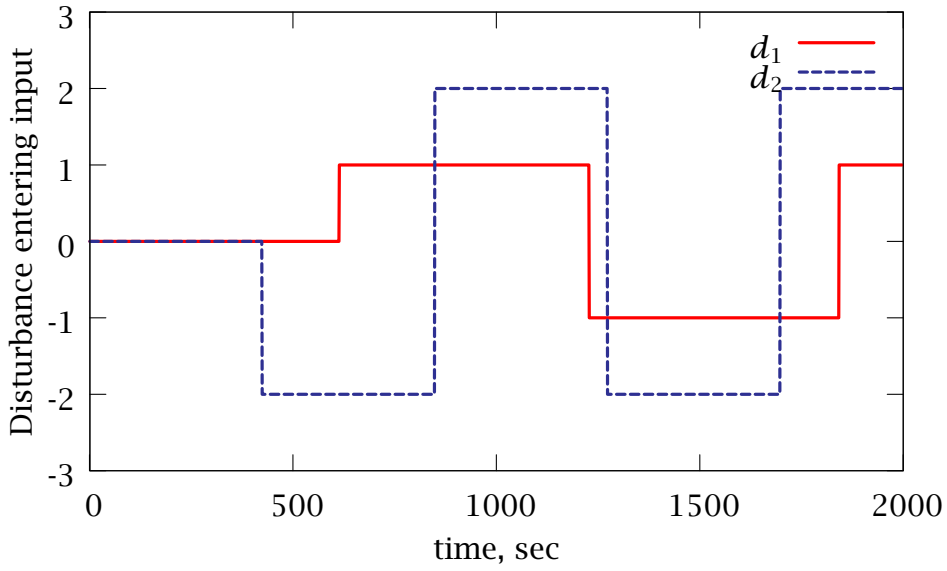


Figure 7.12: Disturbance entering the input in the plant

are:

$$Q_w = \begin{bmatrix} 10^{-5} & 10^{-8} & 10^{-7} & 10^{-10} \\ 10^{-8} & 10^{-4} & 0 & 0 \\ 10^{-7} & 0 & 10^{-5} & 10^{-8} \\ 10^{-10} & 0 & 10^{-8} & 10^{-6} \end{bmatrix} \quad R_v = \begin{bmatrix} 10^{-5} & 0 \\ 0 & 10^{-5} \end{bmatrix}$$

Using the closed-loop data from the plant, with the estimator specified with arbitrary covariances (initial filter gain L_i) , the autocovariance least-squares (ALS) technique is used to find the new Kalman filter gain (say, L_o) for the output disturbance model. Note that in SMOC the disturbance is modelled as entering partly in the input and partly in the output.

If the model had the SMOC structure, then a transformation can be calculated (provided condition given in Chapter 5 are satisfied) from the output disturbance model to this SMOC model, such that the corresponding filter gain for the SMOC model can be calculated knowing L_o. This SMOC model with estimator gain: L_s then gives the same closed-loop performance (same inputs and outputs) as the output disturbance model with L_o

For the particular model in the example, the transformation matrix is uniquely calculated as:

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1.5 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

and the corresponding SMOC filter gain is then $L_s = TL_o$.

For the purpose of comparison, the L estimated from data using the ALS technique for the SMOC model is presented in Table 7.1. The performances for the three cases

Initial tuning L_i	Output model tuning L_o from data	SMOC model $L_s = TL_o$	SMOC model L from data
$\begin{bmatrix} 2 & 0 \\ 0 & 81.84 \\ 0 & 66.77 \\ 0 & 1.5 \\ 2.542 & 0 \\ 0 & 59.76 \end{bmatrix}$	$\begin{bmatrix} -2.950 & 0.027 \\ 1.223 & -144.3 \\ -0.131 & 18.56 \\ -3.512 & -21.11 \\ 2.470 & -0.039 \\ 2.338 & 15.03 \end{bmatrix}$	$\begin{bmatrix} -2.950 & 0.027 \\ 1.223 & -144.3 \\ -0.131 & 18.56 \\ 0.193 & -21.17 \\ 2.470 & -0.039 \\ -0.132 & 15.07 \end{bmatrix}$	$\begin{bmatrix} -2.989 & 0.134 \\ 1.101 & -150.5 \\ -0.107 & 19.00 \\ 0.172 & -22.25 \\ 2.479 & -0.043 \\ -0.110 & 15.70 \end{bmatrix}$

Table 7.1: Estimator gains compared for the SMOC model

(L_i, L_o and L_s) are shown in the Figures 7.13 and 7.14. As seen in Figures 7.13 and 7.14, the inputs and the outputs using the filter gains, L_o and L_s specified using the ALS technique are indistinguishable as expected given the results in Chapter 5. The closed-loop controller performance using the ALS estimator gains is also better compared to the initial estimator gain L_i.

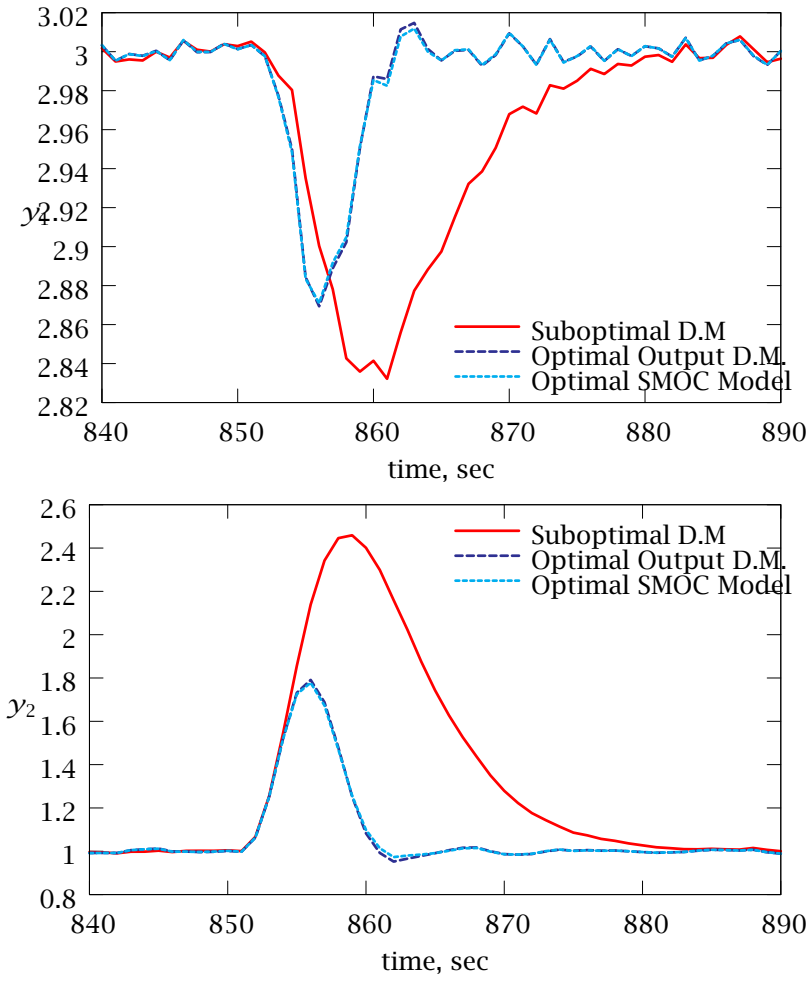


Figure 7.13: Snapshot of the inputs

7.7 Appendix

7.7.1 Plots showing the Improvement in the State Estimates for the Industrial Data Sets I, II, III, IV and V

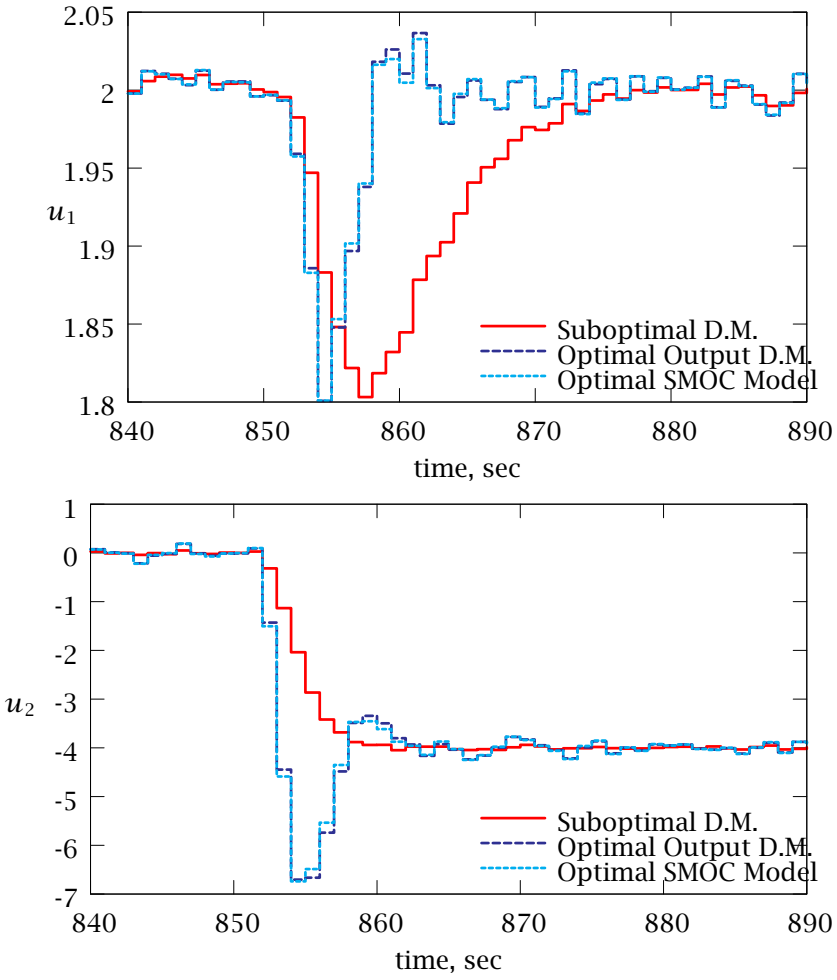


Figure 7.14: Snapshot of the outputs

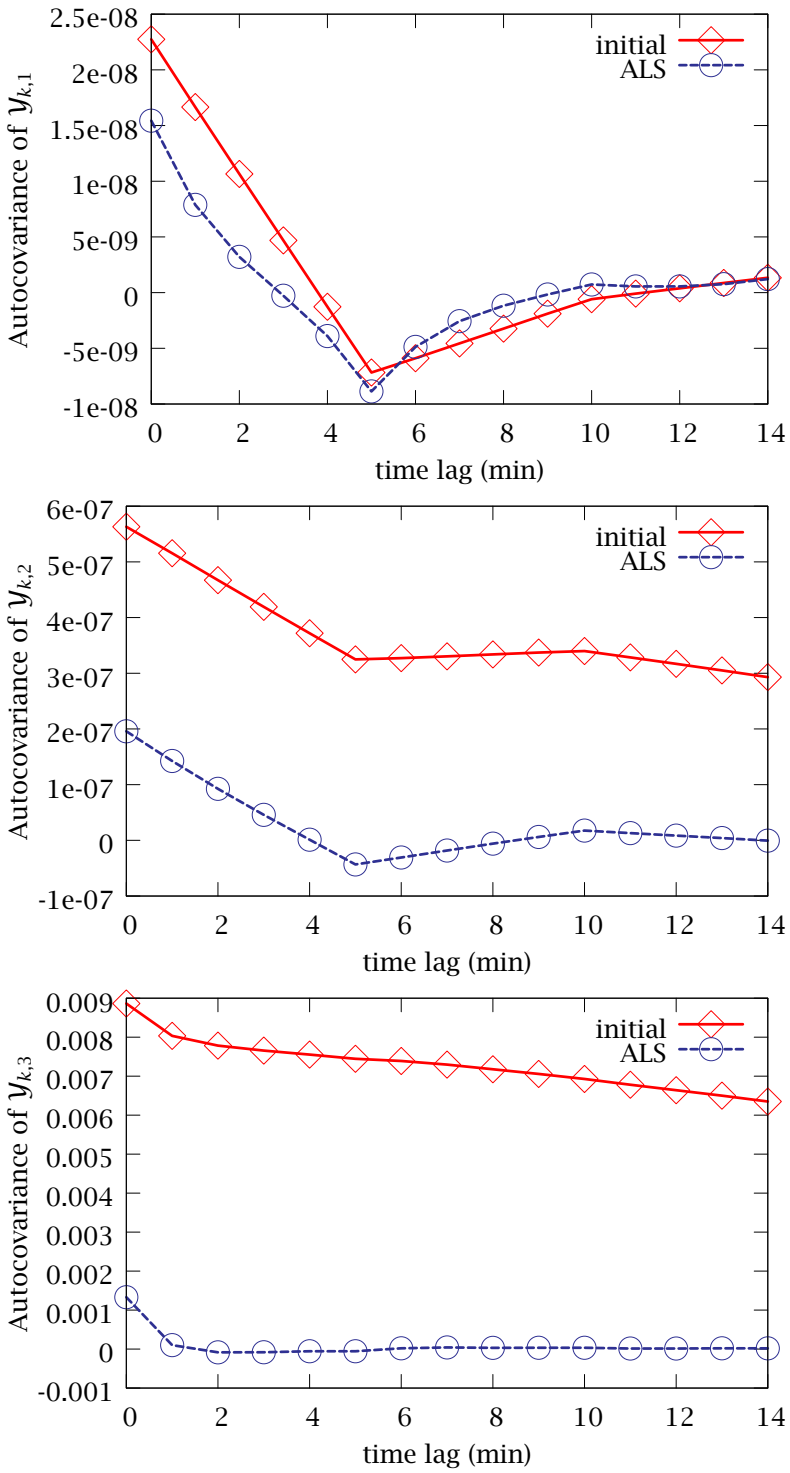


Figure 7.15: Data Set II: Autocovariance of the innovations at different time lags

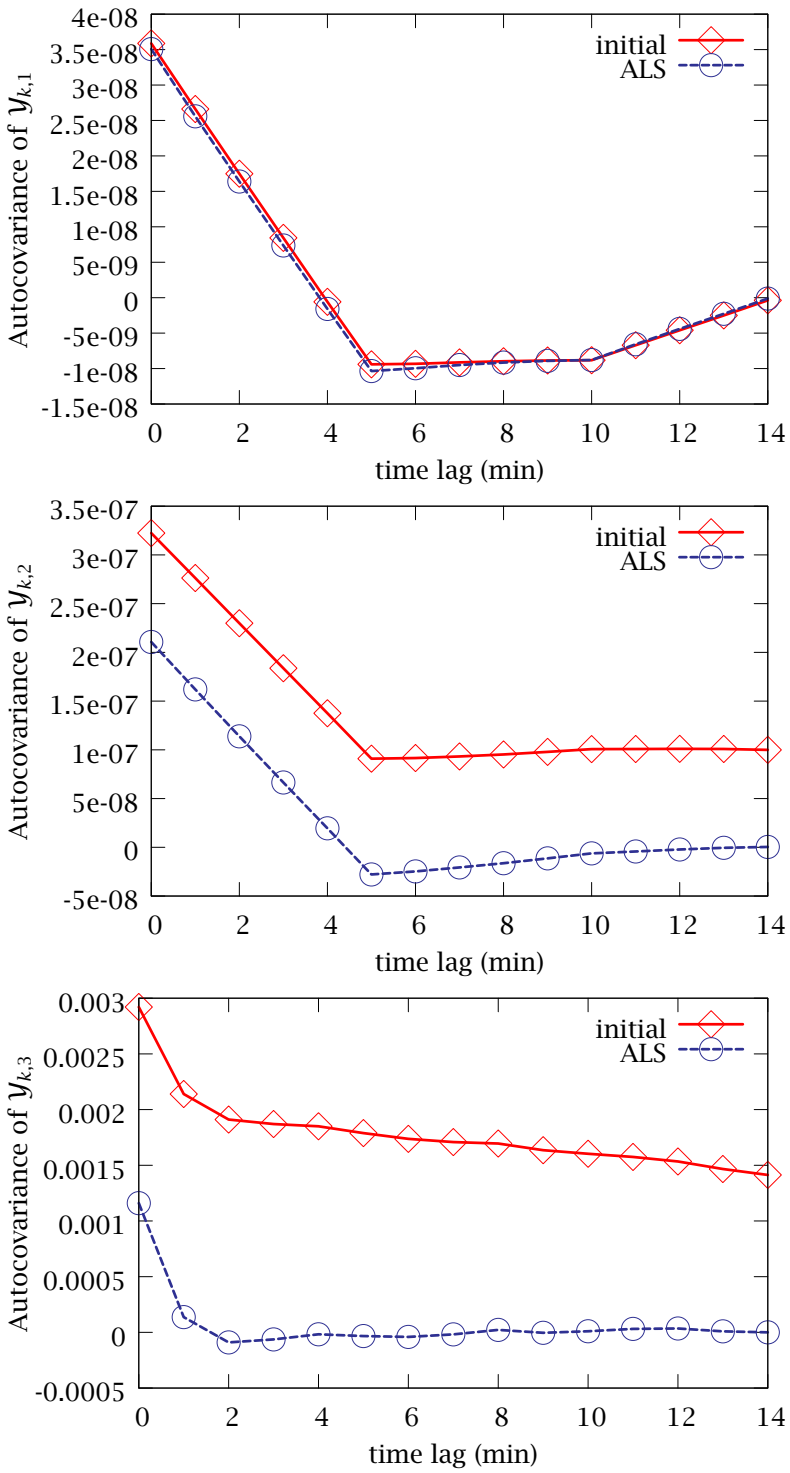


Figure 7.16: Data Set III: Autocovariance of the innovations at different time lags

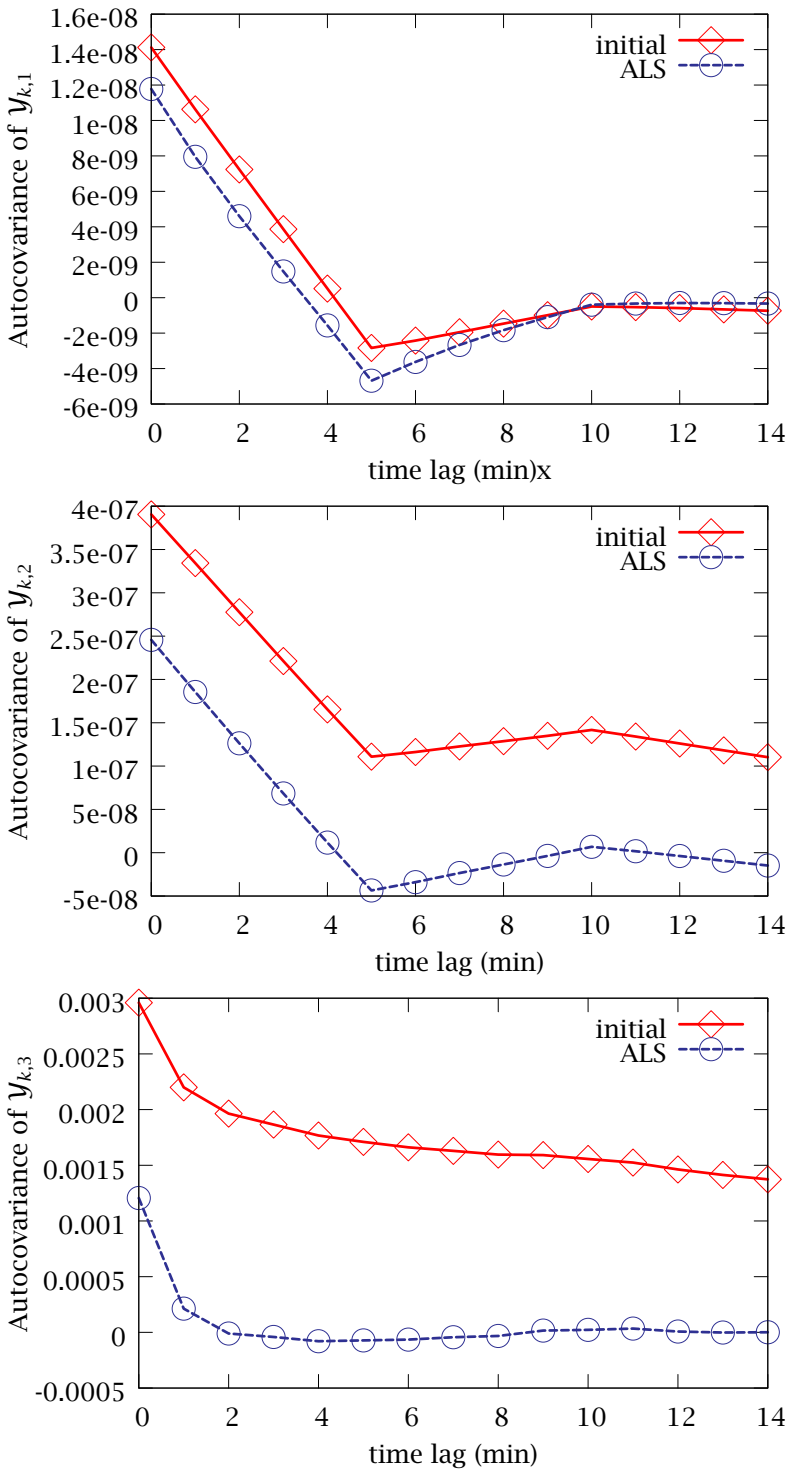


Figure 7.17: Data Set IV: Autocovariance of the innovations at different time lags

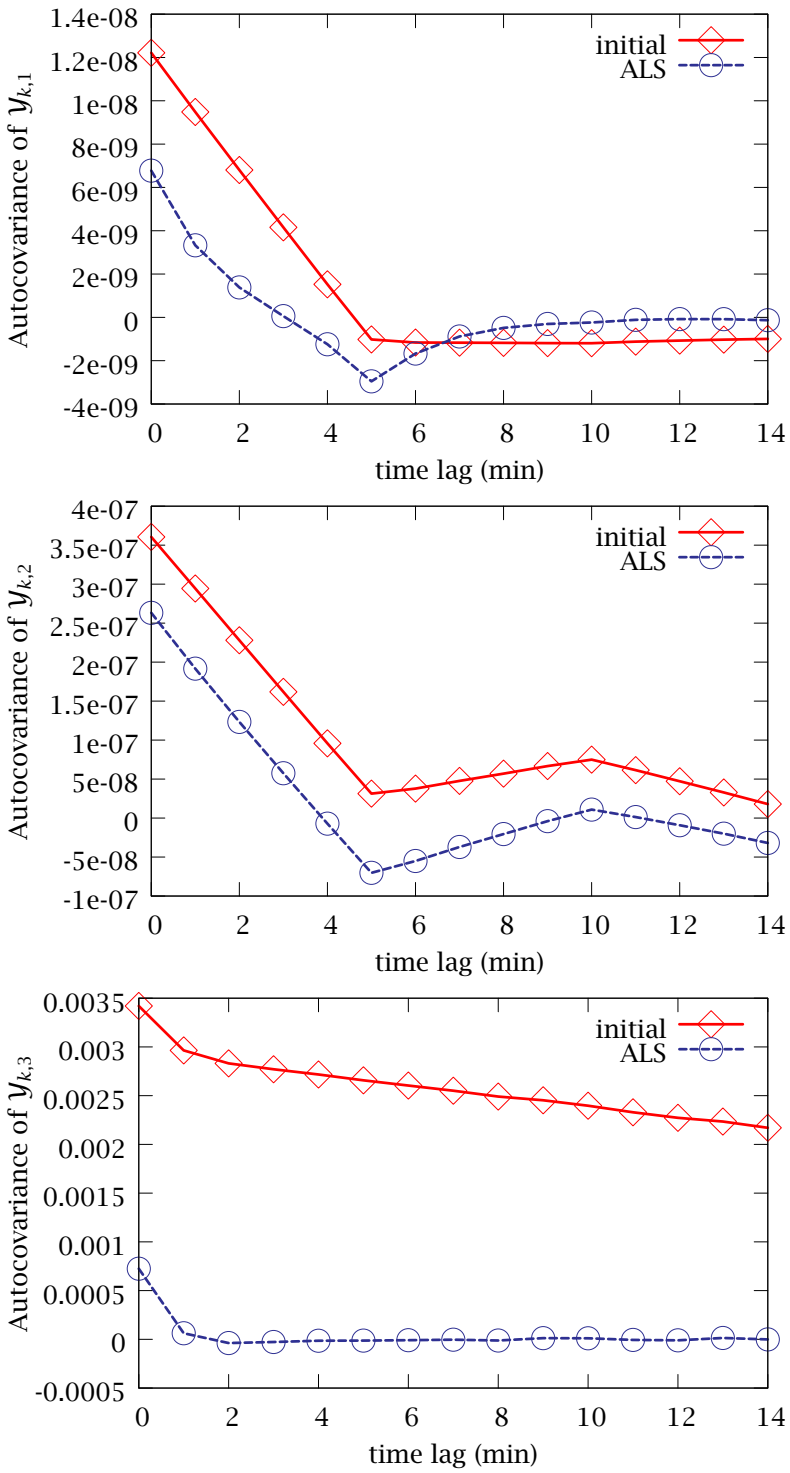


Figure 7.18: Data Set V: Autocovariance of the innovations at different time lags

Chapter 8

Other Nonlinear State Estimation Techniques : Hybrid of Particle Filters and Moving Horizon Estimation [1]

8.1 Introduction

Most processes in the chemical industry exhibit nonlinear dynamics and efficient nonlinear state estimators are an active area of research. The accuracy of the state estimates is important in advanced process control and the performance of a closed-loop controller is directly affected by the quality of the current state estimate. Some state estimators for nonlinear models are Extended Kalman Filters (EKF) (Lee and Ricker, 1994; Myers and Luecke, 1991; Haseltine and Rawlings, 2005; Reif et al., 1999), Moving Horizon Esti-

[1] **Portions of this chapter appear in Rajamani and Rawlings (2007)**

mators (MHE) (Rao et al., 2003; Robertson et al., 1996), Unscented Kalman Filters (UKF) (Julier et al., 2000; Julier and Uhlmann, 2004a,b; Wu et al., 2004) and the Particle Filters (PF) (Doucet et al., 2000; Arulampalam et al., 2002; Rawlings and Bakshi, 2006; Kotecha and Djurić, 2003). Although the EKF has been the most popular in the chemical industry, the performance of the EKF has been criticized in the literature due to the lack of theoretical convergence properties and lack of accuracy in the state estimates when applied to significantly nonlinear models. The unscented Kalman filter is the Kalman filter based on the unscented transformation and was largely pioneered by Julier et al. (2000). The UKF as an alternative to the extended Kalman filter has been studied on several applications eg. Romanenko and Castro (2004); Romanenko et al. (2004); Sitz et al. (2004). The UKF unlike the EKF satisfies convergence properties and the approximation of the model is exact up to the second order. The UKF however, lacks the robustness offered by optimization based state estimators like the moving horizon estimator.

In this chapter, we present a new state hybrid estimator combining the Moving Horizon Estimator (MHE) and the Particle Filter (PF). The advantages of optimizing over a trajectory of states estimates as implemented in the MHE and the computational speed of the PF, combine to give a fast yet robust hybrid estimator. In Section 8.2 state estimation applied to nonlinear models is introduced and the MHE and PF techniques are presented in Sections 8.2.1 and 8.2.2 respectively. We next describe the MHE/PF hybrid technique in Section 8.2.3 and present the results of an application in Section 8.3. Finally, the conclusions are given in Section 8.4.

8.2 Nonlinear State Estimation

Consider the nonlinear discrete-time state-space model given in Chapter 7:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + g(x_k, u_k) w_k \\ y_k &= h(x_k) + v_k \end{aligned} \tag{8.1}$$

where, w_k and v_k are the state and the measurement noises drawn from the Gaussian distributions $N(0, Q_w)$ and $N(0, R_v)$ respectively. The restriction on the noises w_k, v_k being drawn from the Gaussian distribution is not restricting and considered here for simplicity. Given the above system model, at any given time t_k, the measurements $y_{0:k} = \{y_0, y_1 \cdots y_k\}$ and the inputs $u_{0:k} = \{u_0, u_1, \cdots u_k\}$ are available. Here and in the sequel, we use $x_{0:k}$, $u_{0:k}$ and $y_{0:k}$ as shorthand notations for the trajectories of x_k, u_k and y_k respectively. Mathematically, the optimal state estimate $\hat{x}_k$ for the state at time t_k is then defined as:

$$\hat{x}_k = \arg\max_{x_k} p(x_k | y_{0:k}) \tag{8.2}$$

subject to state constraints, the given inputs U_k and the nonlinear model in Equation 8.1.

A closed form analytical solution for Equation 8.2 is usually not available except for the simplest of cases. For the linear model with no constraints and Gaussian distributions for w_k, v_k, Equation 8.2 gives the classical Kalman filter recursion. For nonlinear models, such a simplification is usually not possible. When we have non-Gaussian distributions for w_k, v_k, multi-modality of the state estimates or bad guesses for x_0, the problem of finding the state estimates by solving Equation 8.2 is even more difficult.

8.2.1 Moving Horizon Estimation

Moving horizon estimation (MHE) is an optimization based state estimator, which makes it convenient to handle constraints. In MHE, the optimization variables are the state trajectory over a past horizon of length N rather than just the current state as with EKF, UKF and the PF. The following optimization is then solved as opposed to Equation 8.2:

$$\max_{x_{k-N:k}} p(x_{k-N:k}|y_{0:k}) \tag{8.3}$$

subject to state constraints, the given inputs $u_{0:k}$ and Equation 8.1.

Using the trajectory of states over a past horizon provides robustness against bad data, outliers, bad priors and unmodelled disturbances. If the horizon length is N and the noises w_k, v_k are from the Gaussian distributions $w_k \sim N(0, Q_w)$ and $v_k \sim N(0, R_v)$, the MHE optimization in Equation 8.3 reduces to (Tenny, 2002; Haseltine, 2005):

$$\min_{\hat{x}_{k-N},\cdots\hat{x}_k} (\hat{x}_{k-N} - \bar{x}_{k-N})^T P_{k-N}^{-1} (\hat{x}_{k-N} - \bar{x}_{k-N}) + \sum_{i=0}^{N-1} w_{k-i}^T Q_w^{-1} w_{k-i} + v_{k-i}^T R_v^{-1} v_{k-i}$$

$$\begin{aligned} \text{subject to,} \quad & \hat{x}_{k+1} = f(\hat{x}_k, u_k) + g(\hat{x}_k, u_k) w_k \\ & y_k = h(\hat{x}_k) + v_k \\ & \{\hat{x}_{k-N}, \cdots \hat{x}_k\} \in \mathbb{X} \end{aligned} \tag{8.4}$$

where, $\mathbb{X}$ is the constraint set for the states and $\bar{x}_{k-N}$ and P_{k-N} are the mean and the covariance of the approximated Gaussian distribution of the state x_{k-N}. The performance of the MHE depends on the horizon N and the choice of the approximations for the prior $\bar{x}_{k-N}$ and P_{k-N}. The choice for the weight P_{k-N} is such that it gives a good ap-

proximation of the past data that not included in the horizon. For more details on the choice of $\bar{x}_{k-N}$ and P_{k-N} see Rao et al. (2003). For the unconstrained linear model and N chosen as the full length of data, the optimization in (8.4) simplifies to the Kalman filter equations. The MHE solves an expensive nonlinear optimization at every time step however, and may not be practical to implement in real time for many nonlinear and large-dimensional applications.

8.2.2 Particle Filters

Particle filter is a sampling based approach to state estimation. The ideal sampling based state estimator is one that samples the full posterior probability density $p(x_{0:k}|y_{0:k})$ with large number of samples. If $x_{0:k}^{(i)}$ with $\{i = 1, \cdots, s\}$ denotes s samples of the trajectory $x_{0:k}$ and drawn from the distribution: $p(x_{0:k}^{(i)}|y_{0:k})$, then the marginal $p(x_k|y_{0:k})$ is approximated as the following sampled density:

$$p(x_k|y_{0:k}) \approx \sum_{i=1}^{s} \tilde{w}_k^{(i)} \delta(x_k^{(i)} - x_k) \tag{8.5}$$

where, the weights associated with $x_k^{(i)}$ are chosen as $\tilde{w}_k^{(i)} = 1/s$ and δ represents the Delta function. This sampled density of the marginal from Equation 8.5 can then be used to calculate the current state estimate as:

$$\hat{x}_k = E[x_k|y_{0:k}] = \int x_k p(x_k|y_{0:k}) dx_k \approx \sum_{i=1}^{s} \tilde{w}_k^{(i)} x_k^{(i)} \tag{8.6}$$

However, the full posterior density $p(x_{0:k}|y_{0:k})$ is not available for sampling in most nonlinear cases and the sample and weight pairs cannot be generated as easily as above.

To avoid this issue, samples are drawn from a different distribution called the *Importance Function*, denoted by $\pi(x_{0:k}|y_{0:k})$. The samples $x_{0:k}^{(i)}$ are then associated with the following weights, followed by a normalization step:

$$w_{0:k}^{*(i)} = \frac{p(x_{0:k}^{(i)}|y_{0:k})}{\pi(x_{0:k}^{(i)}|y_{0:k})}, \qquad \tilde{w}_{0:k}^{(i)} = \frac{w_{0:k}^{*(i)}}{\sum_{i=1}^{s} w_{0:k}^{*(i)}} \tag{8.7}$$

The sample and the weight pairs $\{x_{0:k}^{(i)}, \tilde{w}_{0:k}^{(i)}\}$ then represent the sampled density of the probability density $p(x_{0:k}|y_{0:k})$ (Geweke, 1989). In the limit of large samples, the above sampled density converges to $p(x_{0:k}|y_{0:k})$ (Smith and Gelfand, 1992; Gelfand and Smith, 1990). The rate of convergence depends on the choice of the importance function $\pi(x_{0:k}|y_{0:k})$ and how closely it matches the density of interest $p(x_{0:k}|y_{0:k})$.

Since the above sampling procedure is not recursive in time, new samples have to be drawn from the importance function at every time step. To simplify the sampling procedure and to get a recursion in time, the importance function is chosen in the literature, to satisfy the following property (Doucet et al., 2000; Arulampalam et al., 2002):

$$\pi(x_{0:k}|y_{0:k}) = \pi(x_0|y_0) \prod_{i=1}^{k} \pi(x_i|x_{0:i-1}, y_{0:i})$$

The weights in Equation 8.7 can be rewritten using Markov chain properties of the state-space model in Equation 8.1 and using the importance function property above, as the following recursion:

$$w_k^{*(i)} = w_{k-1}^{*(i)} \frac{p(x_k^{(i)}|x_{k-1}^{(i)}) p(y_k|x_k^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k}) p(y_k|y_{0:k-1})}$$

Since the quantity $p(y_k|y_{0:k-1})$ in the denominator above is difficult to evaluate and is independent of the random variables, this quantity is absorbed into the normalization step and the above equation is written as:

$$w_k^{*(i)} = w_{k-1}^{*(i)} \frac{p(x_k^{(i)}|x_{k-1}^{(i)})p(y_k|x_k^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})}, \qquad \tilde{w}_k^{(i)} = \frac{w_k^{*(i)}}{\sum_{i=1}^{s} w_k^{*(i)}}$$

Note that $p(x_k^{(i)}|x_{k-1}^{(i)})$ and $p(y_k|x_k^{(i)})$ are usually available as known functions from Equation 8.1 and can be evaluated.

The particle filter (PF) uses the importance sampling recursion described above and provides state estimates as given in Equation 8.6. The following steps describe the PF algorithm starting from a prior distributed as $x_0 \sim N(\bar{x}_0, P_0)$. The particle filter (also called the sequential importance sampling procedure in the literature) is then implemented with the following steps (Doucet et al., 2000; de Freitas et al., 2000; Arulampalam et al., 2002):

1. Draw s samples $x_0^{(1)}, x_0^{(2)} \cdots x_0^{(s)}$ from the Gaussian prior distribution $x_0 \sim N(\bar{x}_0, P_0)$ and assign weights $w_0^{*(i)} = 1/s$ to each sample.

2. Draw samples for the next time step from the importance function $\pi(x_k^{(i)}|x_{k-1}^{(i)}, y_{0:k})$. For example, if the importance function is chosen as the state propagation Equation 8.1:

$$\pi(x_k^{(i)}|x_{k-1}^{(i)}, y_{0:k}) = f(x_{k-1}^{(i)}, u_{k-1}) + g(x_{k-1}^{(i)}, u_{k-1})w_{k-1}$$

then, s noise samples $m_{k-1}^{(i)}$ are drawn from the normal $N(0, Q_w)$, which is the distribution of the state noise w_{k-1}, and the samples for the next time step are

given as:

$$x_k^{(i)} = f(x_{k-1}^{(i)}, u_{k-1}) + g(x_{k-1}^{(i)}, u_{k-1}) m_{k-1}^{(i)}$$

3. Evaluate the weights associated with the samples $x_k^{(i)}$ up to a normalizing constant using the recursion:

$$w_k^{*(i)} = w_{k-1}^{*(i)} \frac{p(x_k^{(i)}|x_{k-1}^{(i)}) p(y_k|x_k^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})}$$

4. Normalize the weights:

$$\tilde{w}_k^{(i)} = \frac{w_k^{*(i)}}{\sum_{i=1}^{s} w_k^{*(i)}}$$

5. Optionally, resample among the available samples to avoid divergence of the samples and repeat above steps for the next time step. The resampling procedure is described in more details in Smith and Gelfand (1992).

The s sample and weight pairs $\{x_k^{(i)}, \tilde{w}_k^{(i)}\}$ generated by the PF algorithm above are then used to give the state estimates using Equation 8.6 at each time step:

$$\hat{x}_k \approx \sum_{i=1}^{s} w_k^{(i)} x_k^{(i)}$$

The two critical things that dictate the performance of the PF implemented as above are the choice of the importance function and whether and how often resampling occurs. As described in Doucet et al. (2000), the importance function that gives zero variance in the conditional weights is called the optimal importance function and is given by $p(x_k^{(i)}|x_{k-1}^{(i)}, y_k)$ (See Appendix 8.5.2 for a derivation). However, zero variance in

the conditional weights does not imply that the samples track the state perfectly as seen in the example in Appendix 8.5.4. The optimal importance function $p(x_k^{(i)}|x_{k-1}^{(i)}, y_k)$ has a closed form solution whenever the measurement in Equation 8.1 is linear. A more common choice for the importance function is $p(x_k^{(i)}|x_{k-1}^{(i)})$. This is the simplest importance function with the samples propagated through the state equation as seen in the PF algorithm steps described above. Note however, the above importance function does not depend on the measurements and hence the samples diverge whenever the measurement noise covariance is small. Proposition 1 in Doucet et al. (2000) notes that the unconditional variance of the weights increases with time for any choice of the importance function. But as shown in Appendix 8.5.1, the statement proves the increasing variance with time, for a different recursion and not for the particle filter algorithm that is implemented as above.

Resampling is an *ad-hoc* choice that fixes the divergence of the samples (Smith and Gelfand, 1992). The resampling procedure draws samples from the already existing sample locations by interpreting the weights associated with the existing samples as discrete probabilities and resetting the resampled weights to $1/s$. The result of resampling is an increase in the number of samples having higher weights and a decrease in the number samples having smaller weights. As seen in the example in Appendix 8.5.4, the resampling step is required even when the optimal importance function is used to propagate the samples.

Thus, for the particle to perform efficiently, we need a good choice of the im-

portance function and judicious resampling. The advantage of using the particle filter as a state estimator is that the computation is fast and it is easy to code compared to other methods. However, since the PF estimates only the current state at every time instant as opposed to a trajectory of states, it does not provide robustness to unmodelled disturbances as illustrated in the example in Section 8.3.

8.2.3 Combination of MHE with Particle Filters (PF)

To combine the advantages of the MHE and PF techniques, we propose a hybrid MHE/PF technique that works as follows:

- Use the MHE optimization to locate/relocate the samples

- Use the PF to obtain fast state estimates between MHE optimizations

The proposed MHE/PF hybrid has the advantage of fast computation times and robustness to bad priors and unmodelled disturbances. The hybrid estimator offers an advantage whenever the MHE optimization cannot be solved within the sampling time. The hybrid MHE/PF provides a practical and implementable state estimator.

8.3 Example of Improved State Estimation with the MHE/PF hybrid

Consider the following reaction occurring in a CSTR operating under semi-batch mode. Figure 8.1 is a schematic diagram of the process.

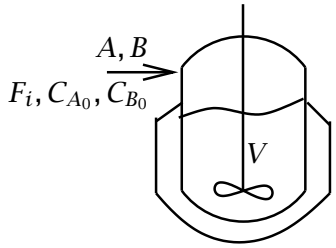


Figure 8.1: Semi-batch CSTR with inflow of A and B

$$2A \xrightarrow{k} B, \qquad k = 0.16 \tag{8.8}$$

The states are the concentrations C_A, C_B and the measurement is the total concentration $(C_A + C_B)$. C_{A_0}, C_{B_0} are the concentrations of A and B in the feed to the reactor. F_i is the flow rate into the reactor and V is the volume. The model for the reactor is then given by:

$$\frac{dC_A}{dt} = -2kC_A^2 + \frac{F_i}{V}C_{A_0}$$
$$\frac{dC_B}{dt} = kC_A^2 + \frac{F_i}{V}C_{B_0}$$

The following values are chosen for the parameters:

$$\frac{F_i}{V} = 0.5, \quad C_{A_0} = 1, \quad C_{B_0} = 0$$

A discretization time of $\Delta t = 0.1$ is chosen and the state $x_{k+1} = \begin{bmatrix} C_A(t_{k+1}) \\ C_B(t_{k+1}) \end{bmatrix}$ is obtained starting from x_k by integrating the continuous time model given above. Gaussian noises w_k, v_k are added to the discretized states and the measurements with covariances $Q_w = \text{diag}\,(0.001^2, 0.001^2)$ and $R_v = 0.1^2$, respectively. The system is simulated with the starting state: $x_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. To compare the performance of the MHE, the PF and the MHE/PF hybrid, we give the bad prior guess of $\bar{x}_0 = [0.1 \quad 4.5]^T$ to the state estimators. The initial covariance is chosen as $P_0 = 10 \times I_2$ to reflect the low confidence in the initial state. An unmodelled disturbance is also added to the states at $t_k = 5$ by pulsing the concentrations C_{A_0}, C_{B_0} in the feed to $[40, 30]^T$ at that time instant.

The MHE described in Section 8.2.1 is implemented with $N = 15, (t = 1.5)$ and a smoothed approximation for the prior $\bar{x}_{k-N}, P_{k-N}$ (Haseltine and Rawlings, 2005). The performance of the MHE is shown in Figure 8.2. The MHE recovers robustly from bad priors and unmodelled disturbances. However, the computation of the MHE implementation is more expensive than that of the PF implementation described next.

Next, the particle filter is implemented with the optimal importance function: $p(x_k|x_{k-1}, y_k)$ (Appendix 8.5.2) with 50 samples and resampling carried out at every time step. Since the measurement equation is linear the optimal importance function has an analytical form and is given in Doucet et al. (2000). The results for the PF implementation are shown in Figure 8.3. As seen in the plot the, PF samples never recover from a bad $\bar{x}_0$, although the measurements are matched accurately.

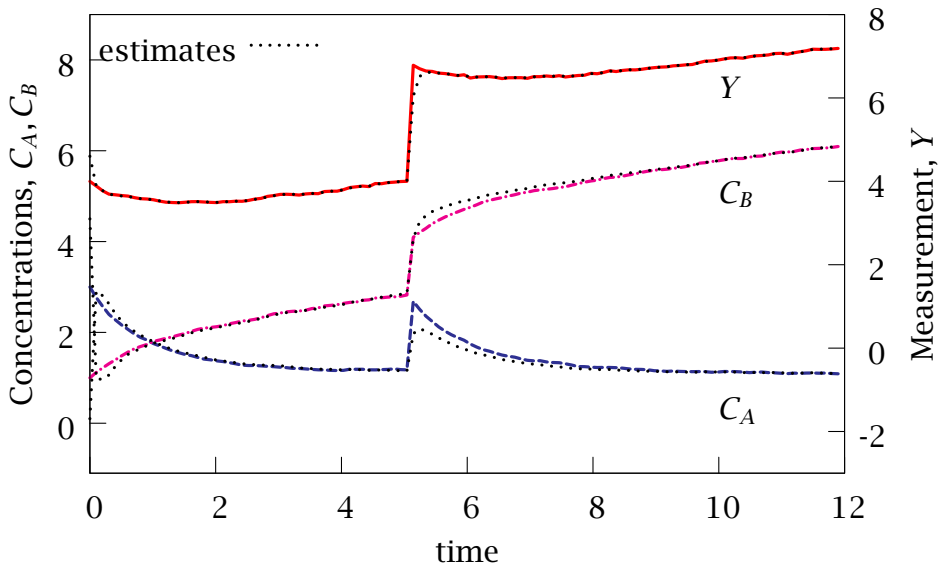


Figure 8.2: MHE with a smoothing prior applied to the semi-batch reactor example

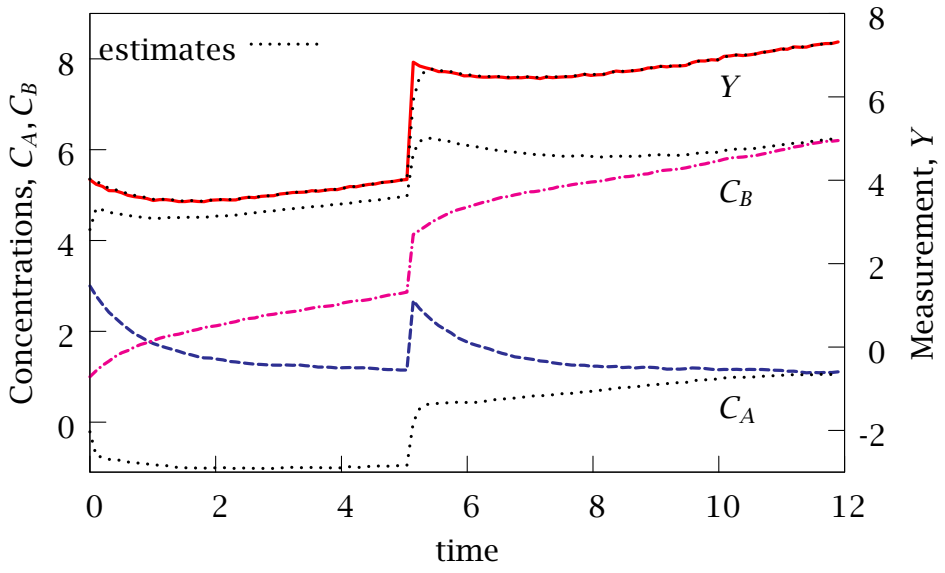


Figure 8.3: PF with the optimal importance function applied to the semi-batch reactor example

We next present the state estimation results from the MHE/PF hybrid implementation with two choices for the importance function. The MHE optimization is carried out at the interval of 15 time steps over a past horizon of $N = 15$. The MHE relocates the samples at these intervals based on a normal distribution with the mean as the MHE smoothed estimate and the smoothing approximation for the covariance. In between the MHE optimizations, the PF samples are used to estimate the states.

Figure 8.4 shows the implementation of the hybrid technique with the importance function chosen as $p(x_k|x_{k-1})$ and with 50 samples. With this simplest importance function, the PF samples recover from a bad $\bar{x}_0$ and the unmodelled disturbance only after the MHE optimization relocates the samples.

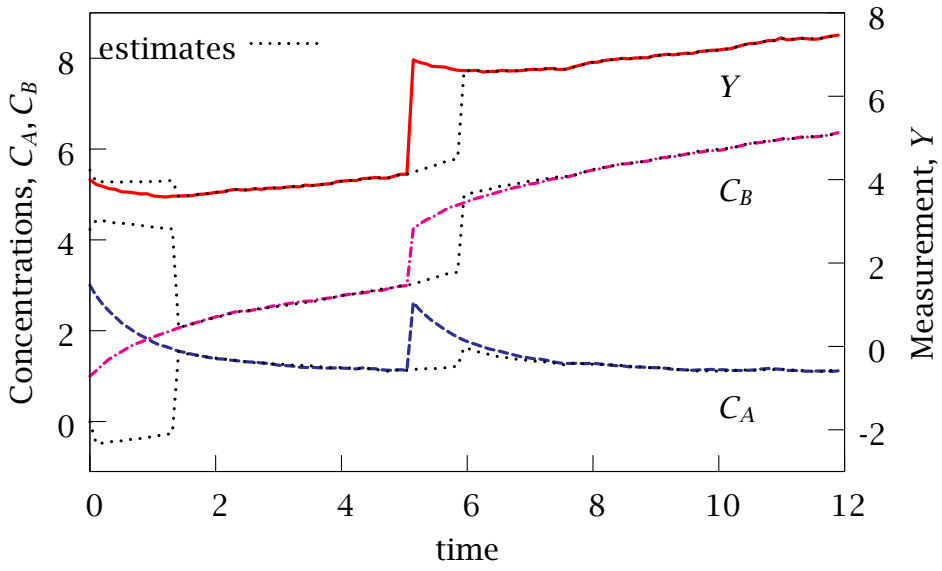


Figure 8.4: MHE/PF hybrid with the simplest importance function applied to the semi-batch reactor example

The performance of the hybrid with the optimal importance function: $p(x_k|x_{k-1}, y_k)$

and 50 samples is more interesting and is shown in Figure 8.5. As seen in the plot, the MHE optimization is required to relocate the samples after a bad $\bar{x}_0$, but the samples recover from the unmodelled disturbance without needing the MHE.

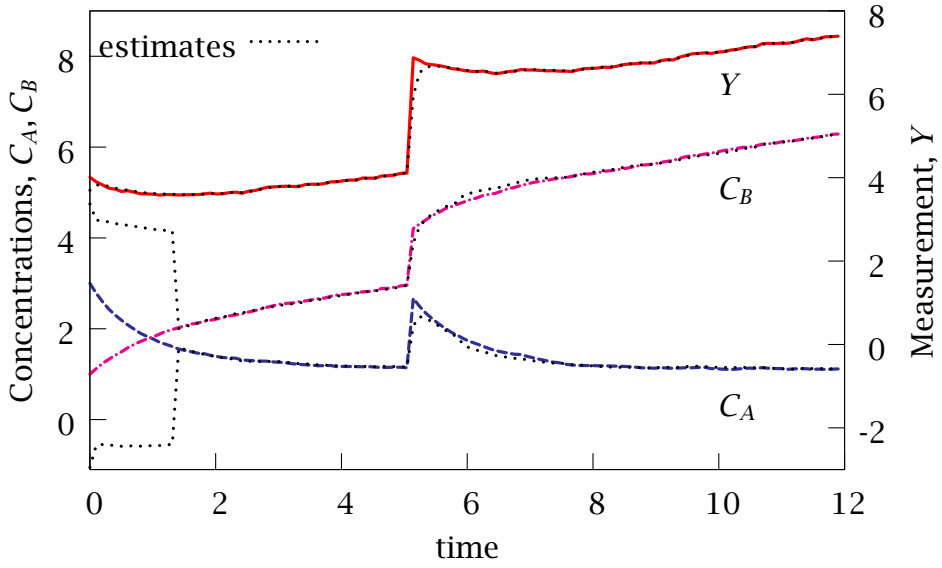


Figure 8.5: MHE/PF hybrid with the optimal importance function applied to the semi-batch reactor example

8.4 Conclusions to the Performance of the MHE/PF hybrid

To compare the performance of the MHE, the PF and the MHE/PF hybrid, we used the following two tests:

Scenario 1: The initial value x_0 is proposed to come from the distribution $N(\bar{x}_0, P_0)$ where $\bar{x}_0$ is incorrect and P_0 is large to account for the low confidence in $\bar{x}_0$. The measurement y_0 is however, close to $h(\bar{x}_0)$ where $h(\cdot)$ is the measurement equation.

Scenario 2: During normal operation, an unmodelled disturbance enters the process at an intermediate time for a short time period. Since the model does not account for this disturbance, the state estimates and the predicted measurements do not reflect the actual system states and measurements at the time instance when the disturbance occurs.

The performance of the MHE, the PF and the MHE/PF hybrid with two choices of the importance function, under the above two Scenarios is described in the following points:

1. MHE is implemented with the smoothed covariance estimate as the arrival cost.

    - Under Scenario 1, MHE reliably fixes bad initial values of $\bar{x}_0$ whenever P_0 is large. A short horizon is enough to recover from the bad initial guess.

    - Under Scenario 2, the performance of the MHE depends on the following factors:

        (a) The characteristics of the disturbance occurring and the covariances chosen for Q_w, R_v.

        (b) The arrival cost approximation. If the arrival cost is too tight as with the smoothed covariance estimate the MHE needs a large horizon to overturn the prior.

        (c) The horizon length N.

2. PF implemented with resampling and with the importance function chosen as $p(x_k|x_{k-1})$.

    - Fails under Scenario 1. The samples do not track either the state or the measurement with bad values of $\bar{x}_0$.

    - Fails under Scenario 2. The samples are unresponsive to unmodelled disturbances and do not track either state or the measurements.

3. PF implemented with resampling and with the optimal importance function $p(x_k|x_{k-1}, y_k)$. Note that a closed form for the optimal importance function is available only when the measurement equation is linear.

    - Can fix Scenario 1 with large number of samples, but is not reliable.

    - Under Scenario 2, performance is similar to the MHE implemented with the smoothed covariance as an approximation for the arrival cost.

4. The PF/MHE hybrid with resampling and the optimal importance function performs better than either technique on its own in terms of accuracy and computation time.

Resampling is required when implementing the PF to prevent sampling of the state-space that do not follow the measurements.

Overall, the MHE has more handles to fix Scenario 2 than the PF but at the cost of additional computation time. The proposed MHE/PF hybrid on the other hand performs

robustly and provides a fix with both Scenarios. The MHE/PF hybrid also provides a reliable state estimator that is implementable in real time. However, The performance of the MHE/PF hybrid depends on the MHE computation time and the discretization time. If the PF is a poor state estimator for a system, then the poor performance of the MHE/PF hybrid scales with the time required to perform a MHE optimization.

8.5 Appendix

8.5.1 Proof of Proposition 1

Here, we prove Proposition 1 given in Doucet et al. (2000), where the claim is made that the unconditional weights in the PF have an increasing variance.

The importance sampling weight recursion equation in the particle filter algorithm is given as:

$$\mathcal{W}_k^{*(i)} = \mathcal{W}_{k-1}^{*(i)} \frac{p(y_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})} \tag{8.9}$$

The unnormalized importance weight is defined as:

$$\mathcal{W}_k^{*(i)} = \frac{p(x_{0:k}^{(i)}|y_{0:k})}{\pi(x_{0:k}^{(i)}|y_{0:k})} \tag{8.10}$$

If $\pi(x_{0:k}^{(i)}|y_{0:k})$ is chosen to satisfy:

$$\pi(x_{0:k}^{(i)}|y_{0:k}) = \pi(x_{0:k-1}^{(i)}|y_{0:k-1})\pi(x_k^{(i)}|y_{0:k}, x_{0:k-1}^{(i)}) \tag{8.11}$$

and the recursion on $p(x_{0:k}^{(i)}|y_{0:k})$ gives (using Markov chain properties):

$$p(x_{0:k}^{(i)}|y_{0:k}) = \frac{p(x_{0:k-1}^{(i)}|y_{0:k-1})p(y_k|x_k^{(i)}p(x_k^{(i)}|x_{k-1}^{(i)})}{p(y_k|y_{0:k-1})} \tag{8.12}$$

We then get,

$$\mathcal{W}_k^{*(i)} = \mathcal{W}_{k-1}^{*(i)} \frac{p(y_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})p(y_k|y_{0:k-1})} \tag{8.13}$$

Note that this is not the same as the recursion given in Equation 8.9. The $p(y_k|y_{0:k-1})$ term in the denominator is ignored because it is assumed to be a constant.

The random variables in this case are $y_{0:k}$ and $x_k^{(i)}$, where $x_k^{(i)}$ are samples drawn from the importance function π.

Since y_k is independent of the samples $x_{0:k-1}^{(i)}$, the joint conditional probability is given by:

$$p(y_k, x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k-1}) = p(y_k|x_{0:k-1}^{(i)}, y_{0:k-1})\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})$$

$$p(y_k, x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k-1}) = p(y_k|y_{0:k-1})\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k}) \tag{8.14}$$

The conditional expectation of $w_k^{*(i)}$ is then given by (with $y_k, x_k^{(i)}$ as random variables):

$$E[w_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k-1}] =$$
$$\int\int w_k^{*(i)} p(y_k, x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k-1}) dy_k dx_k^{(i)} =$$
$$\int\int \frac{w_{k-1}^{*(i)} p(y_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k}) p(y_k|y_{0:k-1})} \pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k}) p(y_k|y_{0:k-1}) dy_k dx_k^{(i)}$$

where we substitute the modified formula for $w_k^{*(i)}$ from Equation 8.13 and Equation 8.14.

We then get,

$$E[w_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k-1}] = w_{k-1}^{*(i)} \int\int p(y_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)}) dx_k^{(i)} dy_k \tag{8.15}$$

and $\boxed{E[w_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k-1}] = w_{k-1}^{*(i)}}$ i.e. $w_k^{*(i)}$ is a martingale. The above proof is similar to the one presented in Kong et al. (1994). For the above martingale, we have:

$$\text{var}\,(w_{k-1}^{*(i)}) = \text{var}\,[E(w_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k-1})]$$
$$= \text{var}\,(w_k^{*(i)}) - E[\text{var}\,(w_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k-1})]$$

The second equality above follows from the law of total variance. See Appendix 8.5.3 for a proof of the law of total variance.

8.5.2 Proof of Proposition 2

Here, we prove Proposition 2 given in Doucet et al. (2000), where the claim is made that the choice $p(x_k|x_{k-1}, y_k)$ for the importance function makes the conditional variance in the weights, zero.

From Equation 8.9,

$$w_k^{*(i)} = w_{k-1}^{*(i)} \frac{p(y_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})}$$

Given a known sequence $\{x_{0:k-1}^{(i)}, y_{0:k}\}$, the random variable $x_k^{(i)}$ is drawn from $\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})$ called the importance function.

Equation 8.9 defines a new random variable $w_k^{*(i)}$ a function of $x_k^{(i)}$. From the definition of conditional expectation we have:

$$E[f(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})] = \int f(x_k^{(i)}) \pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k}) dx_k^{(i)} \tag{8.16}$$

Thus for the random variable $w_k^{*(i)}$, we have:

$$E(w_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k}) = \int \frac{w_{k-1}^{*(i)} p(y_k|x_k^{(i)}) p(x_K^{(i)}|x_{k-1}^{(i)}) \pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})} dx_k^{(i)} \tag{8.17}$$

Also since $w_{k-1}^{*(i)}$ is a constant for a given sequence $(x_{0:k-1}^{(i)}, y_{0:k})$ we get:

$$E(w_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k}) = w_{k-1}^{*(i)} p(y_k|x_{k-1}^{(i)}) \tag{8.18}$$

The conditional variance is then given by:

$$\begin{aligned}\text{var}\ (\mathcal{W}_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k}) =& E[(\mathcal{W}_k^{*(i)} - \mathcal{W}_{k-1}^{*(i)} p(y_k|x_{k-1}^{(i)})^2 | x_{0:k-1}^{(i)}, y_{0:k})] \\ =& E[(\mathcal{W}_k^{*(i)})^2 | x_{0:k-1}^{(i)}, y_{0:k}] - 2E[\mathcal{W}_{k-1}^{*(i)} \mathcal{W}_k^{*(i)} p(y_k|x_{k-1}) | x_{0:k-1}^{(i)}, y_{0:k}] \\ &+ E[(\mathcal{W}_{k-1}^{*(i)})^2 p^2(y_k|x_{k-1}^{(i)}) | x_{0:k-1}^{(i)}, y_{0:k}] \\ =& (\mathcal{W}_{k-1}^{*(i)})^2 \int \frac{p^2(y_k|x_k) p^2(x_k|x_{k-1}^{(i)})}{\pi^2(x_k|x_{0:k-1}, y_{0:k})} \pi(x_k|x_{0:k-1}, y_{0:k}) dx_k \\ &- 2\mathcal{W}_{k-1}^{*(i)} p(y_k|x_{k-1}) E(\mathcal{W}_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k}) + (\mathcal{W}_{k-1}^{*(i)})^2 p^2(y_k|x_{k-1}^{(i)})\end{aligned} \tag{8.19}$$

Thus,

$$\text{var}\ (\mathcal{W}_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k}) = (\mathcal{W}_{k-1}^{*(i)})^2 \left[\int \frac{(p(y_k|x_k) p(x_k|x_{k-1}^{(i)}))^2}{\pi(x_k|x_{0:k-1}^{(i)}, y_{0:k})} dx_k - p^2(y_k|x_{k-1}^{(i)}) \right] \tag{8.20}$$

If $\pi(x_k|x_{0:k-1}^{(i)}, y_{0:k}) = p(x_k|x_{k-1}^{(i)}, y_k) = \frac{p(y_k|x_k)p(x_k|x_{k-1}^{(i)})}{p(y_k|x_{k-1}^{(i)})}$, substituting we get:

$$\text{var}\ (\mathcal{W}_k^{*(i)}|x_{0:k-1}^{(i)}, y_{0:k}) = 0 \tag{8.21}$$

Thus, Proposition 2 is proved.

8.5.3 Proof of Law of Total Variance

For the random variables A and B,

$$\begin{aligned}
\text{var}\,[E_{A|B}(A|B)] &= E_B[E^2_{A|B}(A|B)] - [E_B E_{A|B}(A|B)]^2 \\
&= E_B[E^2_{A|B}(A|B)] - E^2_A(A) \\
&= E_B[E^2_{A|B}(A|B)] - E_A(A^2) + E_A(A^2) - E^2_A(A) \\
&= E_B[E^2_{A|B}(A|B)] - E_B E_{A|B}(A^2|B) + E_A(A^2) - E^2_A(A) \\
&= E_B[E^2_{A|B}(A|B) - E_{A|B}(A^2|B)] + \text{var}\,(A)
\end{aligned}$$

From the above we get the following statement for the law of total variance:

$$\text{var}\,[E_{A|B}(A|B)] = \text{var}\,(A) - E_B[\text{var}\,(A|B)] \tag{8.22}$$

A corollary of the above statement used in the proof of Proposition 1 in Appendix 8.5.1 is:

$$\text{var}\,[E_{A|B}(A|B)] \leq \text{var}\,(A) \tag{8.23}$$

which follows by noting that $E_B[\text{var}\,(A|B)] \geq 0$.

8.5.4 Optimal and the Simplest Importance Function applied to Linear Model

This section compares the application of the particle filter (PF) to a linear model with and without resampling and for two choices of the importance function.

Consider the linear model $A = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$ with $\theta = 6$, $C = \begin{bmatrix} 0.5 & 0.25 \end{bmatrix}$,

$B = I$ and $G = I$. The covariances are chosen as:

$$P_0 = \begin{bmatrix} 1.75 & 1.25 \\ 1.25 & 1.75 \end{bmatrix}, \quad Q_w = 0.01 \times I_2, \quad R_v = 0.01$$

The following inputs are provided to the model simulation for the first five time steps:

$$u_1 = [7, 2]^T, \quad u_2 = [5, 5]^T, \quad u_3 = u_4 = [-1, 2]^T, \quad u_5 = [1, -3]^T$$

We first choose the simplest importance function given by $p(x_k|x_{k-1})$. For the linear model this implies the samples are propagated using the state transition equation by drawing samples from the distribution $N(0, Q_w)$. Note that the propagation of the samples does not depend on the measurements. Figures 8.6 and 8.7 show the propagation of the samples with and without resampling with the simplest importance function. As seen in the plots, the samples start diverging with time and do not track the conditional densities accurately. Resampling fixes the divergence and the brings the samples back to more reasonable locations.

The same simulation is repeated with the optimal importance function used to propagate the samples. Figures 8.8 and 8.9 show the propagation of the samples with and without resampling. Note that the optimal importance function is the probability density $p(x_k|x_{k-1}, y_k)$ and has an analytical solution when the measurement equation is linear. For the linear model the optimal importance function can still give bad performance as seen in Figure 8.8, where the samples do not track the conditional densities accurately. The samples however, do not diverge as with the simplest importance function. Resampling again provides a fix and relocates the samples to more meaningful

locations, as seen in Figure 8.9.

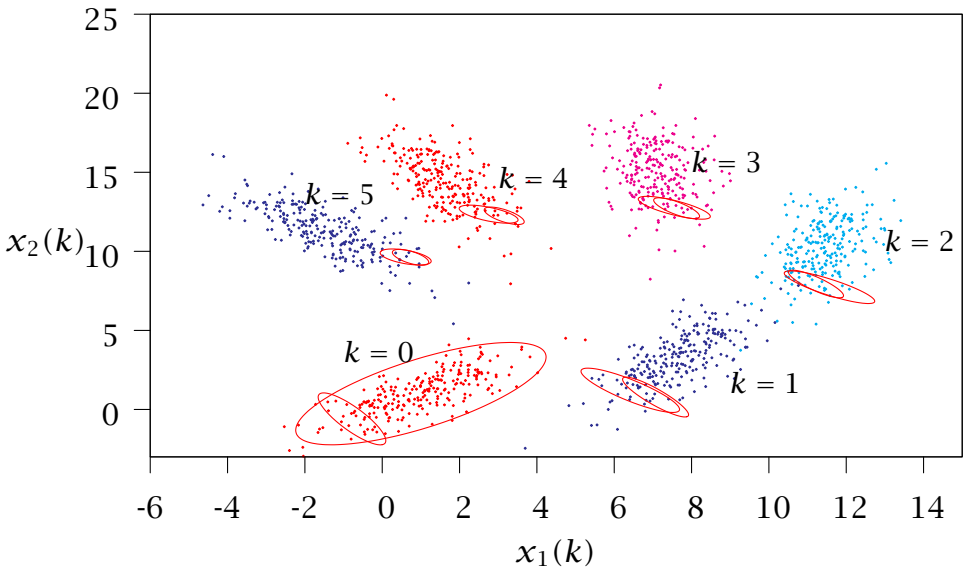


Figure 8.6: Propagation of the samples with the simplest importance function and without resampling. Ellipses represent 95% confidence intervals for true conditional covariances

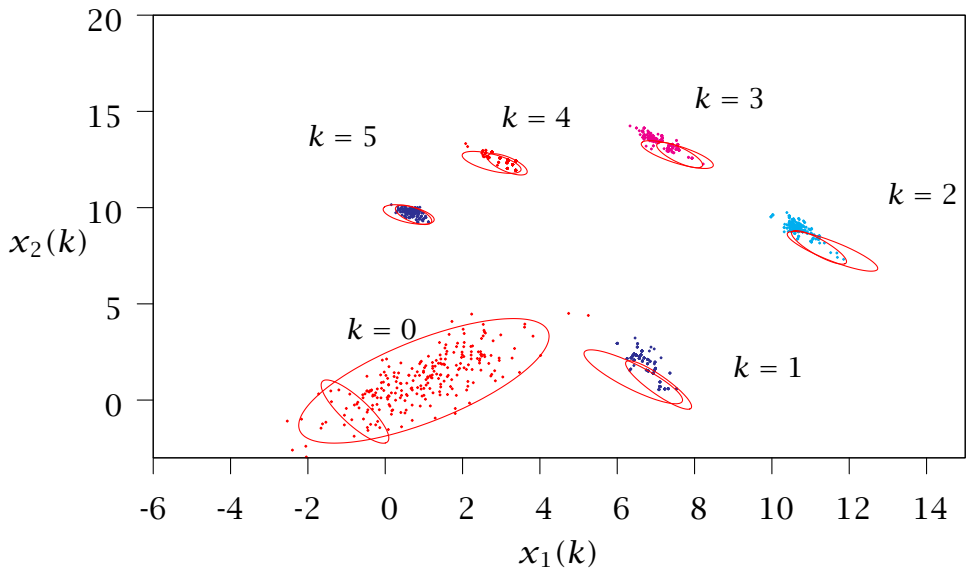


Figure 8.7: Propagation of the samples with the simplest importance function and with resampling. Ellipses represent 95% confidence intervals for the true conditional covariances

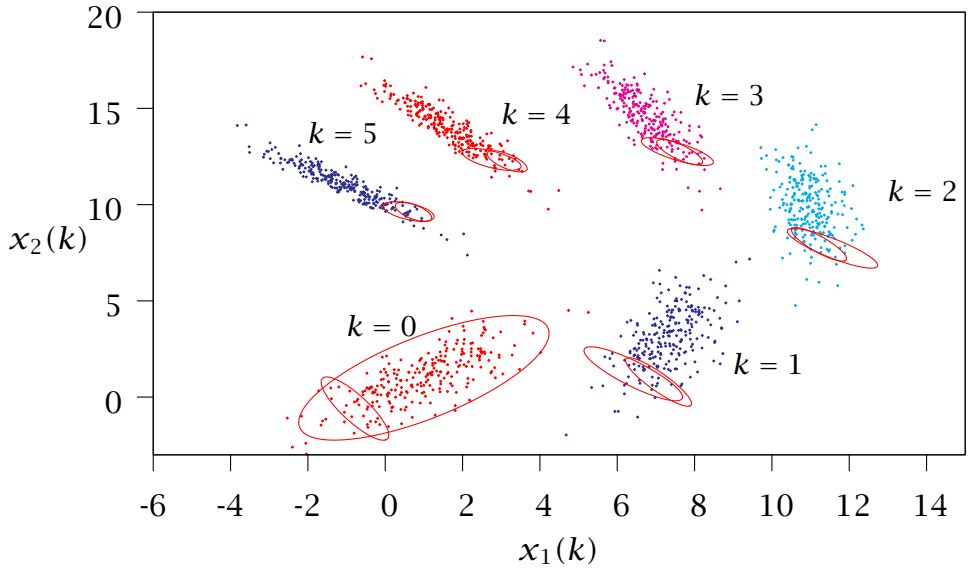


Figure 8.8: Propagation of the samples with the optimal importance function and without resampling. Ellipses represent 95% confidence intervals for the true conditional covariances

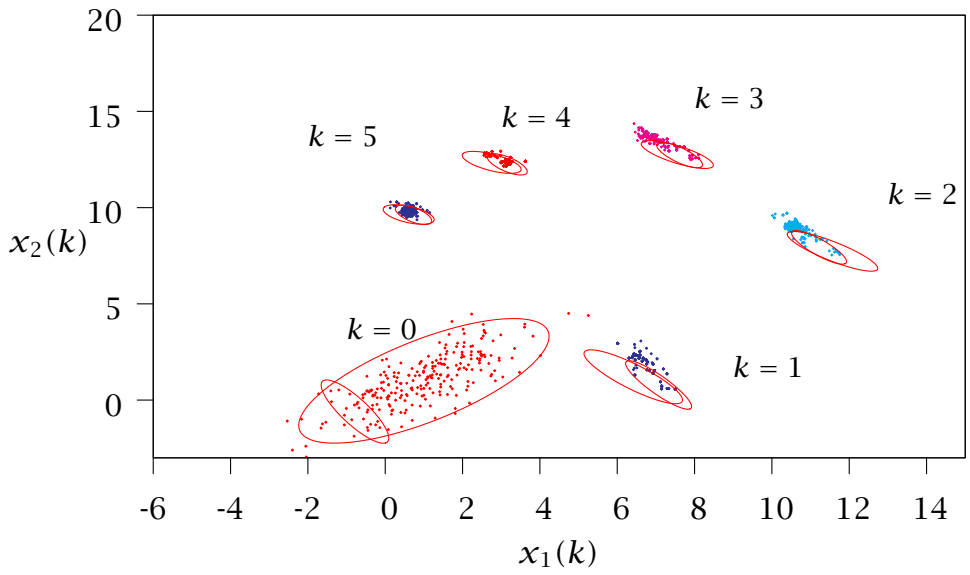


Figure 8.9: Propagation of the samples with the optimal importance function and with resampling. Ellipses represent 95% confidence intervals for the true conditional covariances

Chapter 9

Conclusions and Future Work

This dissertation contains contributions in the area of state estimation applied to model predictive control. The techniques developed in this work are practical and can be easily included in control vendor software to improve the overall control performance. Theoretical results developed in this work provide mathematical support to the developed techniques. Section 9.1 summarizes the major conclusions in this dissertation and Section 9.2 presents proposals for future work in this field.

9.1 Summary of Contributions

- The Autocovariance Least-Squares (ALS) technique was developed based on autocovariances in data to estimate noise statistics in the plant. The ALS technique was shown to give smaller variance in the estimates compared to other correlation based techniques. Simple necessary and sufficient conditions for uniqueness of the ALS estimates were also proved.

- The ALS-SDP technique, that uses the trace heuristic to estimate the minimum number of independent disturbances affecting the data, was developed.

- The optimal weighting for the ALS objective was formulated and examples were presented to illustrate the resulting improvement in the estimates. Other practical implementation issues were presented and simplifications described to help practitioners in coding of the ALS technique.

- The ALS technique was further developed and shown to work with misassigned disturbance models to give offset-free model predictive control (MPC). The explicit connections between ALS and maximum likelihood estimation techniques were also presented.

- Simple modifications to the ALS technique were made to extend the technique to handle nonlinear models and applications to industrial data were shown. Existing control techniques in the industry were also shown to benefit from using the ALS technique to specify the estimator gains.

- Finally a hybrid MHE/PF technique was presented to give good nonlinear state estimation. The hybrid technique uses the optimization approach in MHE to relocate the PF samples robustly while using the PF to provide fast state estimates in between the MHE optimizations. Thus, the hybrid MHE/PF technique combines the complimentary benefits of both the techniques.

9.2 Recommended Future Work

In the following sections possible directions for future work are identified.

9.2.1 Identifying Incorrect Input to Output Model

Model predictive Control (MPC) is a model based control technique and the reliability of MPC methods is only as good as the model. Any errors in the modelling step need to be addressed in an efficient way. For example, using a linear model for a chemical reactor with complex dynamics would induce serious model errors if the operating point is changed. We would require the model to be re-identified whenever the set point changes, in order to get good control performance using MPC. There are various formulations to quantify model errors and one such formulation is robust control, which solves a min-max problem to calculate the optimal inputs for the worst case model error scenario. One could argue that absorbing the model errors as a part of covariance estimation from data would also give reasonably robust performance. The integrating disturbance model discussed in Chapter 5 is a practical technique for accounting for model errors. More research, however, is required to do a comparison of the integrating disturbance model *vis-à-vis* other model error formulations.

A possible approach is to monitor the covariances identified using the ALS technique over time and to check the model for inconsistencies and errors, whenever large changes in the covariance are detected. When there are no model errors the innovations y_k is uncorrelated to the inputs u_k. To detect model errors, correlations between y_k

and u_k can be identified from data. A high correlation value signifies that the errors in the model outputs depend on the corresponding inputs. For example, a high correlation between y_k^1 and u_k^2 would signify that the model relation between the second input and the first output needs to be re-identified.

9.2.2 Quantifying Closed-loop Controller Performance

MPC uses a state estimator and a controller. For the linear time-invariant case with no constraints, the optimal state estimator is the Kalman filter and the optimal controller is the Linear Quadratic Regulator (LQR). The separation principle (Chen, 1999) is then applicable and the overall optimality is due to the individual optimality of the estimator and the controller. For systems that are nonlinear, have constraints and/or have model errors, the separation principle is no longer valid. The state estimator and the controller are then coupled and any change in the estimator tuning requires the controller to be re-tuned to achieve optimality. A possible future direction would be to do a theoretical study to quantify the benefits of the control due to improved state estimation. The theory of optimal control and estimation (Bryson and Ho, 1975; Anderson and Moore, 1979) would be needed to develop the required results. This work would lend mathematical support to the numerical results showing improvement of closed-loop performance using the ALS technique (eg. Chapter 5 and Odelson et al. (2006a)).

9.2.3 Improving Nonlinear State Estimation

The results presented in Chapter 8 provide promise on application of hybrid nonlinear estimation techniques to chemical engineering problems. The development of good importance functions to use in particle filtering is still an active area of research. The use of particle filtering on a trajectory of states as with moving horizon estimation is recommended as future work. A major uncertainty is the lack of knowledge of the noise statistics. A possible future direction is to study the effect of incorrect covariances on the performance of sampling based and hybrid state estimation techniques.

Bibliography

A. C. Aitken. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1935.

B. M. Åkesson, J. B. Jørgensen, N. K. Poulsen, and S. B. Jørgensen. A generalized autocovariance least-squares method for Kalman filter tuning. Accepted for publication in *Journal of Process Control*, June 2007.

B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J., 1979.

B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear Quadratic Methods*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1990. ISBN 0-13-638560-5.

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, third edition, 2003.

M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, February 2002.

K. J. Åström. *Introduction to Stochastic Control Theory*. Academic Press, San Diego, California, 1970.

S. Barnett. Simplification of the lyapunov matrix equation $A^T PA - P = -Q$. *IEEE Transactions on Automatic Control*, 19(4):446–447, 1974.

P. Bélanger. Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, 10:267–275, 1974.

T. Bohlin and S. F. Graebe. Issues in nonlinear stochastic grey box identification. *International Journal of Adaptive Control and Signal Processing*, 9:465–490, 1995.

G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, California, 1st edition, 1970.

G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, Oakland, California, 1976.

G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison–Wesley, Reading, Massachusetts, 1st edition, 1973.

G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, New Jersey, 3rd edition, 1994.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Control Theory*. SIAM, Philadelphia, 1994.

A. E. Bryson and Y. Ho. *Applied Optimal Control*. Hemisphere Publishing, New York, 1975.

B. Carew and P. Bélanger. Identification of optimum filter steady-state gain for systems with unknown noise covariances. *IEEE Transactions on Automatic Control*, 18(6):582–587, 1973.

C. Chen and J. Huang. Estimation of steady-state optimal filter gain from nonoptimal Kalman filter residuals. *Journal of Dynamic Systems, Measurement, and Control*, 116 (3):550–553, September 1994.

C.-T. Chen. *Linear System Theory and Design*. Oxford University Press, 3rd edition, 1999.

C. R. Cutler and B. L. Ramaker. Dynamic matrix control—a computer control algorithm. In *Proceedings of the Joint Automatic Control Conference*, 1980.

E. J. Davison and I. J. Ferguson. The design of controllers for the multivariable robust servomechanism problem using parameter optimization methods. *IEEE Transactions on Automatic Control*, 26(1):93–110, February 1981.

E. J. Davison and A. Goldenberg. Robust control of a general servomechanism problem: the servo compensator. *Automatica*, 11:461–471, 1975.

E. J. Davison and H. W. Smith. Pole assignment in linear time-invariant multivariable systems with constant disturbances. *Automatica*, 7:489–498, 1971.

J. F. G. de Freitas, M. Noranjan, A. H. Gee, and A. Doucet. Sequential Monte Carlo methods to train neural network models. *Neural Computation*, 12:955–993, 2000.

D. Dee, S. Cohn, A. Dalcher, and M. Ghil. An efficient algorithm for estimating noise covariances in distributed systems. *IEEE Transactions on Automatic Control*, 30(11): 1057–1065, 1985.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1–38, 1977.

A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.

D. Dougherty and D. J. Cooper. Tuning guidelines of a dynamic matrix controller for integrating (non-self-regulating) processes. *Industrial and Engineering Chemistry Research*, 42(8):1739–1752, 2003.

J. J. Downs. Control strategy design using model predictive control. In *Proceedings of the American Control Conference, San Diego, CA*, 1999.

S. Eppler. Steady offset and integral control in model predictive control. *Advanced Independent Study*, University of Wisconsin – Madison, 1997.

M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Dept. of Elec. Eng., Stanford University, 2002.

B. A. Francis and W. M. Wonham. The internal model principle for linear multivariable regulators. *J. Appl. Math. Optim.*, 2:170–195, 1975.

B. A. Francis and W. M. Wonham. The internal model principle of control theory. *Automatica*, 12:457–465, 1976.

C. E. García and A. M. Morshedi. Quadratic programming solution of dynamic matrix control (QDMC). *Chemical Engineering Communications*, 46:73–87, 1986.

A. Gelb, editor. *Applied Optimal Estimation*. The M.I.T. Press, Cambridge, Massachusetts, 1974.

A. Gelfand and A. Smith. Sampling based approaches to calculating marginal densities. *Journal of the Americal Statistical Association*, 85:398–408, 1990.

M. Gevers. A personal view of the development of system identification. *IEEE Control Systems Magazine*, 26(6):93–105, December 2006.

J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57(6):1317–1339, November 1989.

G. A. Ghazal and H. Neudecker. On second-order and fourth-order moments of jointly distributed random matrices: a survey. *Linear Algebra and its Applications*, 321:61–93, 2000.

G. C. Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

A. Graham. *Kronecker products and matrix calculus with applications*. Ellis Horwood Limited, West Sussex, England, 1981.

Y. Gupta. Control of integrating processes using dynamic matrix control. *Chemical Engineering Research and Design*, 76(A4):465–470, May 1998.

N. B. Haaser and J. A. Sullivan. *Real Analysis*. Dover Publications, Inc., New York, 1991.

S. Hammarling. Numerical solution of the stable, non-negative definite lyapunov equation. *IMA Journal of Numerical Analysis*, 2:303–323, 1982.

E. L. Haseltine. *Systems Analysis of Stochastic and Population Balance Models for Chemically Reacting Systems*. PhD thesis, University of Wisconsin–Madison, 2005.

E. L. Haseltine and J. B. Rawlings. Critical evaluation of extended Kalman filtering and moving horizon estimation. *Industrial and Engineering Chemistry Research*, 44(8): 2451–2460, April 2005. URL `http://pubs.acs.org/journals/iecred/`.

J. Hedengren, K. Allsford, and J. Ramlal. Moving horizon estimation and control for an industrial gas phase polymerization reactor. In *American Control Conference*, New York, USA, July 11-13 2007.

B. Ho and R. Kalman. Efficient construction of linear state variable models from input/output functions. *Regelungstechnik*, 14:545–548, 1966.

S. Hoang, P. De Mey, and O. Talagrand. A simple adaptive algorithm of stochastic approximation type for system parameter and state estimation. In *Proceedings of the 33rd*

Conference on Decision and Control, pages 747–752, Lake Buena Vista, FL, December 1994.

S. Hoang, P. De Mey, O. Talagrand, and R. Baraille. A new reduced-order adaptive filter for state estimation in high-dimensional systems. *Automatica*, 33(8):1475–1498, 1997.

D. Hua. On the symmetric solutions of linear matrix equations. *Linear Algebra and its Applications*, 131:1–7, 1990.

A. Isaksson. Identification of time varying systems through adaptive Kalman filtering. In *Proceedings of the 10th IFAC World Congress, Munich, Germany*, pages 305–310. Pergamon Press, Oxford, UK, 1987.

G. Jenkins and D. Watts. *Spectral Analysis and its Applications*. Holden-Day, 500 Sansome Street, San Fransisco, California, 1968.

J. Juang. *Applied System Identification*. Prentice Hall, Englewood Cliffs, New Jersey, 1994.

J. Juang and M. Phan. Identification of system, observer, and controller from closed-loop experimental data. *Journal of Guidance, Control, and Dynamics*, 17(1):91–96, January-February 1994.

J. Juang, C. Chen, and M. Phan. Estimation of Kalman filter gain from output residuals. *Journal of Guidance, Control, and Dynamics*, 16(5):903–908, September-October 1993.

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

S. J. Julier and J. K. Uhlmann. Corrections to unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(12):1958, December 2004b.

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

R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Trans. ASME, J. Basic Engineering*, pages 95–108, March 1961.

G. Kitagawa. An algorithm for solving the matrix equation $X = FXF^T + S$. *International Journal of Control*, 25(5):745–753, 1977.

A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the Americal Statistical Association*, 89(425):278–288, March 1994.

J. H. Kotecha and P. M. Djurić. Gaussian particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2592–2601, October 2003.

N. R. Kristensen, H. Madsen, and S. B. Jørgensen. Parameter estimation in stochastic grey-box models. *Automatica*, 40:225–237, 2004a.

N. R. Kristensen, H. Madsen, and S. B. Jørgensen. A method for systematic improvement of stochastic grey-box models. *Computers & Chemical Engineering*, 28:1431–1449, 2004b.

W. Larimore. Automated and optimal system identification by canonical variables. In *Proceedings of the American Control Conference*, pages 1640–1644, Baltimore, Maryland, June 1994.

W. Larimore. Statistical optimality and canonical variate analysis system identification. *Signal Processing*, 52(2):131–144, 1996.

J. Lee and N. Ricker. Extended Kalman filter based nonlinear model predictive control. *Industrial and Engineering Chemistry Research*, 33:1530–1541, 1994.

J. H. Lee, M. Morari, and C. E. García. State-space interpretation of model predictive control. *Automatica*, 30(4):707–717, 1994.

S. Li, K. Y. Lim, and D. G. Fisher. A state space formulation of model predictive control. *AIChE Journal*, 35(2):241–249, February 1989.

W. Li and S. Qin. Consistent dynamic PCA based on error-in-variables subspace identification. *Journal of Process Control*, 11(11):661–678, 2001.

L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

P. Lundström, J. H. Lee, M. Morari, and S. Skogestad. Limitations of dynamic matrix control. *Computers & Chemical Engineering*, 19:409–421, 1995.

J. R. Magnus and H. Neudecker. The commutation matrix: Some properties and applications. *Annals of Statistics*, 7(2):381–394, March 1979.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, New York, 1999.

P. Marquis and J. P. Broustail. SMOC, a bridge between state space and model predictive controllers: Application to the automation of a hydrotreating unit. In T. J. McAvoy, Y. Arkun, and E. Zafiriou, editors, *Proceedings of the 1988 IFAC Workshop on Model Based Process Control*, pages 37–43, Oxford, 1988. Pergamon Press.

J. L. Maryak, J. C. Spall, and B. D. Heydon. Use of the Kalman filter for inference in state-space models with unknown noise distributions. *IEEE Transactions on Automatic Control*, 49(1):87–90, January 2004.

R. Mehra. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 15(12):175–184, 1970.

R. Mehra. On-line identification of linear dynamic system with application to Kalman filtering. *IEEE Transactions on Automatic Control*, 16(1):12–21, 1971.

R. Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 17: 903–908, 1972.

M. Morari and J. H. Lee. Model predictive control: The good, the bad, and the ugly. In Y. Arkun and W. H. Ray, editors, *Chemical Process Control—CPC IV*, pages 419–444, Amsterdam, 1991. Fourth International Conference on Chemical Process Control, Elsevier.

M. Morari and G. Stephanopoulos. Part III: Optimal selection of secondary measurements within the framework of state estimation in the presence of persistent unknown disturbances. *AIChE Journal*, 26(2):247–260, 1980.

K. R. Muske and T. A. Badgwell. Disturbance modeling for offset-free linear model predictive control. *Journal of Process Control*, 12(5):617–632, 2002.

K. R. Muske and J. B. Rawlings. Model predictive control with linear models. *AIChE Journal*, 39(2):262–287, 1993a.

K. R. Muske and J. B. Rawlings. Linear model predictive control of unstable processes. *Journal of Process Control*, 3(2):85–96, 1993b.

M. Myers and R. H. Luecke. Process control applications of an extended Kalman filter algorithm. *Computers & Chemical Engineering*, 15(12):853–857, 1991.

C. Neethling and P. Young. Comments on “Identification of optimum filter steady-state gain for systems with unknown noise covariances”. *IEEE Transactions on Automatic Control*, 19(5):623–625, 1974.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.

G. Noriega and S. Pasupathy. Adaptive estimation of noise covariance matrices in real-time preprocessing of geophysical data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5):1146–1159, 1997.

B. J. Odelson. *Estimating Disturbance Covariances From Data for Improved Control Performance*. PhD thesis, University of Wisconsin–Madison, 2003.

B. J. Odelson, A. Lutz, and J. B. Rawlings. The autocovariance least-squares methods for estimating covariances: Application to model-based control of chemical reactors. *IEEE Control Systems Technology*, 14(3):532–541, May 2006a.

B. J. Odelson, M. R. Rajamani, and J. B. Rawlings. A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308, February 2006b. URL `http://www.elsevier.com/locate/automatica`.

M. Palaniswami and G. Feng. Adaptive control algorithms for disturbance rejection. *Computers and Electrical Engineering*, 17(1):31–37, 1991.

G. Pannocchia. Robust offset-free model predictive control. IFAC 2002 World Congress, Barcelona, Spain, 2002.

G. Pannocchia. Robust disturbance modeling for model predictive control with application to multivariable ill-conditioned processes. *Journal of Process Control*, 13(8): 693–701, 2003.

G. Pannocchia and A. Bemporad. Combined design of disturbance model and observer

for offset-free model predictive control. *IEEE Transactions on Automatic Control*, 52 (6):1048–1053, June 2007.

G. Pannocchia and J. B. Rawlings. Disturbance models for offset-free MPC control. *AIChE Journal*, 49(2):426–437, 2002.

S. J. Qin and T. A. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7):733–764, 2003.

S. J. Qin, W. Lin, and L. Ljung. A novel subspace identification approach with enforced causal models. *Automatica*, 41(12):2043–2053, December 2005.

M. R. Rajamani and J. B. Rawlings. Estimation of noise covariances and disturbance structure from data using least squares with optimal weighting. In *Proceedings of AIChE Annual Meeting*, San Francisco, California, November 2006.

M. R. Rajamani and J. B. Rawlings. Improved state estimation using a combination of moving horizon estimator and particle filters. In *Proceedings of the American Control Conference*, New York, USA, July 11-13 2007.

M. R. Rajamani and J. B. Rawlings. Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. Accepted for publication in *Automatica*, September, 2007.

M. R. Rajamani, J. B. Rawlings, S. J. Qin, and J. J. Downs. Equivalence of MPC disturbance

models identified from data. In *Proceedings of Chemical Process Control—CPC 7*, Lake Louise, Alberta, Canada, January 2006.

M. R. Rajamani, J. B. Rawlings, and S. J. Qin. Achieving state estimation equivalence for misassigned disturbances in offset-free model predictive control. Submitted for publication in *AIChE Journal*, August, 2007.

M. R. Rajamani, J. B. Rawlings, and T. A. Soderstrom. Application of a new data-based covariance estimation technique to a nonlinear industrial blending drum. Submitted for publication in *IEEE Transactions on Control Systems Technology*, October, 2007.

C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained linear state estimation – a moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.

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

J. B. Rawlings and B. R. Bakshi. Particle filtering and moving horizon estimation. *Computers & Chemical Engineering*, 30:1529–1541, 2006.

J. B. Rawlings and K. R. Muske. Stability of constrained receding horizon control. *IEEE Transactions on Automatic Control*, 38(10):1512–1516, October 1993.

K. Reif, S. Günther, E. Yaz, and R. Unbehauen. Stochastic stability of the discrete-time

extended Kalman filter. *IEEE Transactions on Automatic Control*, 44(4):714–728, April 1999.

J. Richalet, A. Rault, J. L. Testud, and J. Papon. Model predictive heuristic control: Applications to industrial processes. *Automatica*, 14:413–428, 1978.

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

A. Romanenko and J. A. Castro. The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study. *Computers & Chemical Engineering*, 28(3):347–355, March 15 2004.

A. Romanenko, L. O. Santos, and P. A. F. N. A. Afonso. Unscented Kalman filtering of a simulated pH system. *Industrial and Engineering Chemistry Research*, 43:7531–7538, 2004.

F. C. Schweppe. Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, 11(1):61–70, January 1965.

P. O. Scokaert, J. B. Rawlings, and E. S. Meadows. Discrete-time stability with perturbations: Application to model predictive control. *Automatica*, 33(3):463–470, 1997.

F. G. Shinskey. *Feedback Controllers for the Process Industries*. McGraw-Hill, New York, 1994.

R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the em algorithm. *Journal of Time Series Analysis*, 3:253–264, 1982.

R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer-Verlag, New York, 2000.

A. Sitz, U. Schwarz, and J. Kurths. The unscented Kalman filter, a powerful tool for data analysis. 14(6):2093–2105, 2004.

A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2):84–88, 1992.

A. G. Sparks and D. S. Bernstein. Optimal rejection of stochastic and deterministic disturbances. *Journal of Dynamic Systems, Measurement, and Control*, 119:140–143, March 1997.

W. Steeb. *Kronecker product of matrices and applications*. Mannheim, 1991.

D. S. Stoffer and K. Wall. Bootstrapping state space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the Americal Statistical Association*, 86:1024–1033, 1991.

R. K. Sundaram. *A First Course in Optimization Theory*. Cambridge University Press, 1996.

M. Tenny. *Computational Strategies for Nonlinear Model Predictive Control*. PhD thesis, University of Wisconsin–Madison, 2002.

M. J. Tenny, J. B. Rawlings, and S. J. Wright. Closed-loop behavior of nonlinear model predictive control. *AIChE Journal*, 50(9):2142–2154, September 2004. URL `http://www.aiche.org/aichejournal/`.

C. F. Van Loan. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics*, 123:85–100, 2000.

P. Van Overschee and B. De Moor. N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.

P. Van Overschee and B. De Moor. A unifying theorem for three subspace system identification algorithms. *Automatica*, 31(12):1853–1864, December 1995.

J. VanAntwerp and R. Braatz. A tutorial on linear and bilinear matrix inequalities. *Journal of Process Control*, 10(4):363–385, 2000.

L. Vandenberghe and S. Boyd. A primal-dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 69(1):205–236, 1995.

L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.

M. Viberg. Subspace methods for the identification of linear time-invariant systems. *Automatica*, 31(12):1835–1851, 1995.

H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*. Kluwer Academic Publishers, Massachusetts, 2000.

Y. Wu, M. Wu, D. Hu, and X. Hu. An improvement to unscented transformation. *Lecture Notes in Artificial Intelligence*, 3339:1024–1029, 2004.

N. Xu and T. Shou. Optimal moving average filter with application to self-tuning type control against deterministic disturbances. *International Journal of Systems Science*, 22(8):1449–1466, 1991.

W. Zhuang, M. R. Rajamani, J. B. Rawlings, and J. Stoustrup. Application of autocovariance least-squares method for model predictive control of hybrid ventilation in livestock stable. In *Proceedings of the American Control Conference*, New York, USA, July 11-13 2007a.

W. Zhuang, M. R. Rajamani, J. B. Rawlings, and J. Stoustrup. Model predictive control of thermal comfort and indoor air quality in livestock stable. In *Proceedings of the European Control Conference*, Kos, Greece, July 2-5 2007b.

Vita

Murali Rajaa Rajamani was born in Coimbatore, India on 6 June, 1980. He grew up and lived in Mumbai, India since the age of three and finished his high school and junior college studies from Swami Vivekanand High School/Junior College, Chembur, Mumbai in 1998. He received his Bachelors in Chemical Engineering (B. Chem. Engg.) degree with a gold medal from Mumbai University - Institute of Chemical Technology (formerly University Department of Chemical Technology, Mumbai University) in 2002. In Fall 2002, he moved to Madison to brave the brutal Wisconsin winters and to start working towards a Ph.D. degree in the Department of Chemical and Biological Engineering, University of Wisconsin-Madison under the supervision of Prof. James B. Rawlings. Murali will begin work in the Advanced Control and Instrumentation group at BP Research and Technology in Naperville, Illinois after graduation.

Permanent Address: Pune, India

This dissertation was prepared with LaTeX 2ε[1] by the author.

[1] This particular University of Wisconsin compliant style was carved from The University of Texas at Austin styles as written by Dinesh Das (LaTeX 2ε), Khe-Sing The (LaTeX), and John Eaton (LaTeX). Knives and chisels wielded by John Campbell and Rock Matthews.