Model predictive control with tracking error bound and an influence function approach to moving horizon estimation

Zhou, Dexiang

2017

Zhou, D. (2017). Model predictive control with tracking error bound and an influence function approach to moving horizon estimation. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/69439>

<https://doi.org/10.32657/10356/69439>

**MODEL PREDICTIVE CONTROL WITH
TRACKING ERROR BOUND AND AN
INFLUENCE FUNCTION APPROACH TO
MOVING HORIZON ESTIMATION**

DEXIANG ZHOU

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2017

To my family

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

Date

DEXIANG ZHOU

Acknowledgement

I would like to begin by expressing my deepest gratitude to my supervisor, Associate Professor Ling Keck Voon, for his guidance, understanding, inspiration, invaluable advice and kindly help, for allowing me considerable freedom in doing this research. He gave me a lot of support on the projects I am interested in, which is a very valuable experience for me. I learned tolerant, careful and logical from him when we discussed papers, which helped me a lot not only with this thesis, but also with the right way to do research.

I would like to thank Associate Professor Ho Weng Khuen from National University of Singapore for discussion on the influence function approach to ARMAX process with t -distribution noise.

I would like to thank Prof. J.M. Maciejowski from Cambridge University for discussion on the influence function approach to ARMAX process with t -distribution noise and symmetry between MPC and MHE.

I would like to thank my friends, Liu Zhichao, Jing jin, Yang Rong, Nikita, Lin Liyong, Ahmad, Yin Le, Chen Tengpeng, Kou Fei, Wu Baoquan, Antonis, Hung, Thuy and so on. I really enjoyed life here with these wonderful guys.

I would like to thank my apartment mates, Weng Renling, Wu Kai, Zhou Zhi and Qi Yongbin. I really enjoy sharing an apartment with them.

At last, but not definitely not least, my family. I could never have achieved this goal without their endless support. Thanks my parents for raising and supporting

me. Special thanks to my wife, Fan Congmin, for accompanying and caring about me along the way.

Summary

Model Predictive Control (MPC) and constrained Moving Horizon Estimation (MHE) are both optimization-based methods where a constrained optimization problem is solved at each time instant. Recursive feasibility of the constrained optimization problem plays a key role in MPC as it ensures that the state is not driven into a "blind alley" where the constrained problem could become infeasible. For MHE the constrained optimization problem can become infeasible when erroneous measurements occur. In this thesis, we showed the relationship between a trajectory advisor for MPC with tracking error bound and a measurement checker for constrained MHE. Akin to the concept of reference governor, the trajectory advisor can supply a new feasible reference signal to recover the feasibility of the MPC problem if the original reference signal makes the MPC problem infeasible. The measurement checker could be used to determine the admissibility of the measurements before the MHE is solved.

MPC requires state feedback and if the full state is not measurable, a state estimator is required. Gaussian noise is often assumed in state estimation and the optimality of the estimators based on this assumption may be lost when the assumption is violated. Constrained MHE offers an approach to handle non Gaussian noise by using constraints. In this thesis, a novel moving horizon Maximum Likelihood Estimation (MLE) is proposed for Autoregressive-Moving-Average with exogenous (ARMAX) process with t -distribution noise model. The thick tail prop-

erty of t -distribution downs weigh outliers in the likelihood function so that the proposed estimator is robust to outliers. Instead of solving the maximization problem numerically, the Influence Function (IF), an analysis tool in robust statistics, is used to formulate a computationally efficient recursive solution to the maximization problem approximately. As expected, the recursive estimator is reduced to the traditional Moving Window Least-Squares Estimation when the noise is Gaussian. The moving horizon N can be used as tuning parameter for the proposed estimator. The formula for the variance of the estimate is derived. This formula shows explicitly how the variance of estimate is affected by the number of measurements and noise variance. In addition, a formula which predicts the effect of an outlier on estimate is derived. The simulations show the advantages of the proposed estimator over Moving Horizon Least-Squares Estimation in the sense of estimation variance and robustness to outlier, and the advantage over Kalman filter when model mismatch happens. The proposed estimator is applied to liquid level estimation of a Coupled-Tank Control Apparatus and indoor environment localization which show that the proposed estimator gives smaller estimation variance and is more robust to outlier than Moving Window Least-Squares Estimation.

Contents

Acknowledgements	i
Summary	iii
List of Contents	v
List of Figures	viii
List of Tables	xii
1 Introduction	1
1.1 Model Predictive Control	1
1.2 Moving Horizon Estimation	6
1.3 Contributions	16
2 Background	19
2.1 Model Predictive Control	19
2.2 Moving Horizon Estimation	26
2.3 Symmetry between control and estimation	27
2.4 Limitation of Kalman filter	28
2.5 Influence Function	29
2.6 ARMAX process	31

3	The Effect of Sample/Hold Time on Initial Feasible Set in Model Predictive Control Design	33
3.1	Introduction	33
3.2	MPC Design Based on Continuous Time Cost Function	35
3.3	Relationship between Initial feasible set of MPC and S/H time	39
3.4	Illustrative Example	43
3.5	Conclusion	48
4	A Trajectory Advisor for Model Predictive Control with Tracking Error Bound and a Measurement Checker for Moving Horizon Estimation	49
4.1	Introduction	50
4.2	Problem Formulation — MPC with tracking error bound	52
4.3	Trajectory Advisor for MPC with tracking error bound	56
4.4	Measurement Checker for MHE	57
4.5	Illustrative Examples	61
4.6	Conclusion	69
5	Moving Horizon Estimation for ARMAX process with t-Distribution Noise	71
5.1	Introduction	72
5.2	Maximum Likelihood Estimation for ARMAX process with t -distribution noise	74
5.3	Moving Horizon Estimation of ARMAX process with t -distribution noise (MHE-TD)	79
5.4	Statistical Interpretation	83
5.5	Simulations	86
5.6	Experiments	94

5.7	Conclusion	102
6	Conclusion and Future Work	104
	Author's Publications	111
	Appendix	124
1	Proof of Theorem 2.1	125
2	From Linear Quadratic Tracking to Kalman Filter by symmetry . . .	126
3	Direct Method to compute the set of admissible measurements of MHE (4.13a)	133
4	Derivation of (5.10)	134
5	Proof of Proposition 5.1	136
6	Proof of Corollary 5.1	138
7	Proof of Proposition 5.2	139
8	Proof of Theorem 5.1	141
9	Proof of Theorem 5.2	144
10	Proof of Corollary 5.3	146

List of Figures

1.1	Initial feasible set of the MPC when $h = 0.2$ s and $h = 0.5$ s.	4
1.2	Estimation performance with $Q = 10 \times I_4$ and $R = 5$	9
1.3	Estimation performance with $Q = 0.1 \times I_4$ and $R = 0.1$	9
1.4	Estimation of Velocities using EKF and constrained MHE when Q and R are chosen correctly.	12
1.5	Estimation of Velocities using EKF and constrained MHE when Q and R are inaccurate.	13
1.6	Doppler Frequency Estimation.	14
1.7	The histogram of measurement noise of indoor environment localization.	15
2.1	Reference Governor.	24
3.1	Stable system with two distinct real roots $p_1 = -6$, $p_2 = -4$: (a) size (area) of initial feasible set vs h ; (b) the initial feasible sets. Note that the initial feasible set with $h = 0.4$ s includes the one with $h = 0.25$ s while the initial feasible set with $h = 0.25$ s doesn't include the one with $h = 0.2$ s.	40
3.2	Terminal invariant sets of $h = 0.5$ s and $h = 1$ s.	41
3.3	Initial feasible sets of $h = 0.5$ s and $h = 1$ s.	41

3.4	Unstable system with two distinct real roots $p_1 = 5$, $p_2 = 3$: (a) size (area) of initial feasible set vs h ; (b) the initial feasible sets. Note that the initial feasible set with $h = 0.2$ s does not include the one with $h = 0.25$ s.	43
3.5	Unstable system with one stable root and one unstable root $p_1 = -6$, $p_2 = 3$: (a) size (area) of initial feasible set vs h ; (b) the initial feasible sets. Note that the size of the initial feasible set are almost the same when the S/H time changes.	44
3.6	(a) Closed-loop cost vs S/H time h ; (b) Area of initial feasible set vs S/H time h	46
3.7	Computational time vs S/H time.	47
3.8	Initial feasible set of the MPC when $h = 0.2$ s, $h = 0.25$ s and $h = 0.4$ s.	47
4.1	Symmetry between the MHE (4.13) and the MPC with tracking error bound (4.14)	60
4.2	Comparison of MPC with and without tracking error bound, output of trajectory advisor y_k^{r*} and the feasible reference set $\Omega_{y_N^r}$ (the vertical red line).	63
4.3	System inputs and incremental inputs of the MPC with and without tracking error bound	63
4.4	With incremental model, comparison of MPC with and without tracking error bound, output of trajectory advisor y_k^{r*} and the feasible reference set $\Omega_{y_N^r}$ (the vertical red line).	65
4.5	System input and incremental input of the MPC with and without tracking error bound.	65
4.6	The system output y_k , reference signal y_k^r and the output of trajectory advisor y_k^{r*}	66
4.7	Tracking error.	67

4.8	The system input.	67
4.9	The set of admissible measurements.	69
5.1	x_T and $\hat{x}_T$ of IF approximation of MHE-TD.	81
5.2	Variance of estimation versus N	89
5.3	Variance of estimation versus noise shape parameter q	89
5.4	Estimation error of IF approximation of MHE-TD, Moving Window Least-Squares estimation, ARMAX filter and Kalaman filter when model mismatch happened at $25 < T < 29$: (a) $N=3$; (b) $N=6$; (c) $N=9$	92
5.5	Simulation results of Simulation 3: (a) $\hat{y}_T$ of IF approximation of MHE-TD; (b) $\hat{y}_T$ of Moving Window Least-Squares Estimation. . . .	94
5.6	Experiment apparatus: Coupled-Tank Control Apparatus.	95
5.7	Measurements of Experiment y_k	95
5.8	Noise distribution and fitting.	96
5.9	Estimation result: Estimation of Liquid Level by using Moving Win- dow Least-Squares Estimation.	97
5.10	Estimation result: Estimation of Liquid Level by using the IF ap- proximation of MHE-TD.	97
5.11	$\hat{y}_k$ of IF approximation of MHE-TD.	98
5.12	$\hat{y}_k$ of Moving Window Least-Squares Estimation.	98
5.13	Indoor environment localization test-bed.	99
5.14	The histogram of noise e_k	99
5.15	The results of the trilateration.	100
5.16	Estimation results of IF approximation of MHE-TD and Moving Win- dow Least-Squares Estimation.	102
5.17	$\hat{y}_k$ of IF approximation of MHE-TD and Moving Window Least Squares Estimation.	102

6.1	(a) MLE estimate as a function of h (solid line) and the tangent at $h = 0$ (dashed line) for t_3 noise; (b) Second order derivative of $\kappa(h)$ wrt. h	106
-----	--	-----

List of Tables

1.1	Performance Comparison of Kalman Filter and MHE	10
4.1	Variables translation (see ref. [76], Table 1)	59
4.2	Parameters translation (see ref. [76], Table 2)	59
5.1	Parameters of the ARMAX process, the IF approximation of MHE- TD and ARMAX filter in the Simulations.	87
5.2	Variance of $\hat{y}_T$ in Simulation 1.	89
5.3	Variance of $\hat{y}_T$ in Simulation 2.	93
5.4	Variance of $\hat{y}_T$ in Figures 5.9 and 5.10	97
1	Variables translation between LQT and Kalman filter	128
2	Parameters translation between LQT and Kalman filter	128

Nomenclature

$(\cdot)'$	transpose operation
$E(\cdot)$	expectation operation
I_n	identity matrix with dimension of $n \times n$
$Var(\cdot)$	variance operation
$\mathbf{0}_{m \times n}$	zero matrix with dimension of $m \times n$
$\text{Re}(\cdot)$	Real part of a complex value
$x_{i,k}$	the i^{th} element of state x at time instant k

Chapter 1

Introduction

1.1 Model Predictive Control

Consider the following system to be controlled

$$x_{k+1} = f(x_k, u_k), \quad (1.1)$$

where $f(\cdot)$ is implicitly defined by the originating differential equation that has an equilibrium point at the origin ($f(0, 0) = 0$). The constraints on control input and state sequences are

$$u_k \in \mathbb{U}, x_k \in \mathbb{X}, \quad (1.2)$$

where $\mathbb{U}$ is a convex, compact subset of $\mathbb{R}^m$ and $\mathbb{X}$ a convex, closed subset of $\mathbb{R}^n$, each set containing the origin in its interior.

Given the current state x_0 , the standard MPC is defined as [78]

$$\begin{aligned} \min_{\{u_k\}_{k=0}^{N-1}} V(x_0, \{u_k\}_{k=0}^{N-1}) &= \sum_{k=0}^{N-1} \ell(x_k, u_k) + F(x_N), \\ s.t. \quad x_{k+1} &= f(x_k, u_k), \quad (k = 0, \dots, N-1) \end{aligned}$$

$$\begin{aligned}
u_k &\in \mathbb{U}, \quad (k = 0, \dots, N-1) \\
x_k &\in \mathbb{X}, \quad (k = 0, \dots, N) \\
x_N &\in \mathbb{X}_f \subseteq \mathbb{X},
\end{aligned} \tag{1.3}$$

where we assume the state cost $\ell(x, u) \geq c(|(x, u)|)^2$ and that $\ell(0, 0) = 0$. x_k denotes the predicted state vector obtained by applying the input sequence $\{u_i\}_{i=0}^{k-1}$ to the system (1.1) starting at x_0 . The terminal constraint $x_N \in \mathbb{X}_f \subseteq \mathbb{X}$ is usually imposed. $F(x_N)$ is the terminal cost. N is the prediction horizon (here the control horizon is also N). The optimal control sequence $\{u_k^*\}_{k=0}^{N-1}$ is obtained by solving the constrained optimization (1.3). The first control (i.e. u_0^*) of the sequence $\{u_k^*\}_{k=0}^{N-1}$ is applied to the plant, in order to generate feedback to possible disturbance. In the next instant the above procedure is repeated.

The feasibility and stability of MPC for linear time invariant systems is well established [60, 78, 97]. In dual-mode prediction paradigm [97], a terminal invariant set is used to ensure recursive feasibility of MPC. This terminal invariant set is the maximal output admissible set [37] of the linear system under a fixed unconstrained feedback controller which is often chosen as the controller of infinite unconstrained Linear Quadratic Regulator (LQR).

The initial feasible set, i.e., the largest subset of the state space for which there exists a controller action satisfying all the constraints, is obtained by calculating backward from the terminal invariant set. The recursive feasibility of MPC is then guaranteed by imposing the initial state of the system to be included in the initial feasible set. The size of the initial feasible set is affected by several factors. One of the factors is the prediction horizon; generally longer horizons result in larger initial feasible set [60]. Another factor is the Sample/Hold (S/H) time. For example,

consider the following continuous time linear system

$$\dot{x}_t = A_c x_t + B_c u_t \quad (1.4)$$

where

$$A_c = \begin{bmatrix} 11 & -30 \\ 1 & 0 \end{bmatrix}, \quad B_c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and t is the continuous time index and the control input u_t is constant over a duration h , i.e.,

$$u_t = u_k, \text{ for } t \in [kh, (k+1)h)$$

where h is the S/H time and k is the sampling instant. The Zero Order Hold (ZOH) discrete time state space model of (1.4) is given by

$$x_{k+1} = Ax_k + Bu_k \quad (1.5)$$

where

$$A = e^{A_c h}, \quad B = \int_0^h e^{A_c \tau} d\tau B_c.$$

Assume that the model (1.5) is subject to the following state and input constraints:

$$\begin{bmatrix} -1 \\ -1 \end{bmatrix} \leq x_k \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad -1 \leq u_k \leq 2.5.$$

The corresponding initial feasible sets with $h = 0.2$ s and $h = 0.5$ s are shown in Figure 1.1 where the size of initial feasible set with $h = 0.2$ s is bigger than $h = 0.5$ s. As far as we know, the effect of S/H time on the initial feasible set is not investigated in the literature. In this thesis, we studied preliminary investigation and observation on this effect. To preliminarily investigate and observe how the size of initial feasible set is affected by MPC design parameters, we formulate MPC

Figure 1.1: Initial feasible set of the MPC when $h = 0.2$ s and $h = 0.5$ s.

based on continuous time cost function and state space model to account for the discretization effect when S/H time changes. The effect of S/H time on the size of initial feasible set is investigated and shown through simulations by employing a second order system. Further we give an example to show the selection of S/H time in MPC design, especially from the perspective of trade-off between their on-line computational complexity, initial feasible set, and closed-loop performance.

Besides regulation, MPC is often required to track a time varying reference, such as piece-wise constant references, random varying reference [30], or a target set (zone) while ensuring that the state and control constraints, and even bound tracking error are satisfied all the times. Changes in the reference signals may cause MPC to become infeasible. For constrained tracking problem, reference governor was originally designed in [40] to guarantee its feasibility. The basic idea of the reference governor is to add an auxiliary nonlinear device between the reference command input and the closed-loop control system. The auxiliary nonlinear device on-line adjusts the trajectory to be followed. Several MPC schemes have been proposed to guarantee the feasibility of constrained tracking problem [70, 71]. In [71] admissible

steady state and input are introduced as additional decision variables in a modified cost function of MPC. The control law ensures feasibility for piecewise constant reference. A robust version of [71] which could handle additive disturbance by using tube-based MPC is investigated in [70]. Another way to tackle the recursive feasibility of the MPC tracking problem is to convert the MPC tracking problem to an equivalent MPC regulation problem [82]. Then the initial feasible set of the equivalent MPC regulation problem indicates the condition of recursive feasibility of the MPC tracking problem. In this thesis, the initial feasible set of a standard MPC with tracking error bound is determined. Based on the feasible set, a trajectory advisor is constructed to recover the feasibility of the MPC by supplying a feasible reference to MPC if the original reference makes the MPC infeasible.

The potential application of the trajectory advisor is dual-stage automated tooling machines [23, 45] where a slow stage with large workspace (also called operating range) moves a stage with much faster dynamics and limited workspace that actuates the tool. The fast stage stroke constraint requires the difference between the target position and the slow stage position to be within the range of the fast stage. Because of the stages timescale separation (often 2-3 orders of magnitude), to ensure tool trajectory tracking it is enough to control the slow stage. The MPC with trajectory advisor can be used to control the slow stage to maintain the difference between the slow stage trajectory and the target trajectory within the range of the fast stage, while enforcing the operating constraints. We showed in simulations the performance of MPC with trajectory advisor. In addition, we highlighted that to achieve offset free tracking in the presence of unknown constant disturbance, incremental model should be used and the cost function suitably modified.

1.2 Moving Horizon Estimation

State estimation is usually required in process control and monitoring, and Gaussian noise is often assumed in state estimation. However, the Gaussian noise assumption is an approximation to reality. Constraints are useful to model the non Gaussian noise, such as truncated Gaussian noise [43] and the mixtures of two Gaussian noises [94]. One class of estimators which can handle constraints is constrained MHE [43, 88, 89]. A key feature of constrained MHE is its ability to take the prior information (e.g., constraints on state and noises) into account which prevents the estimator from giving estimates which are not reasonable (i.e. the values which violate the constraints). MHE solves online optimization problem (such as, constrained convex optimization) based on the measurements within a moving window. Consider the following system

$$x_{k+1} = f(x_k, k) + \omega_k, \quad (1.6a)$$

$$y_k = h(x_k, k) + e_k, \quad (1.6b)$$

where y_k is the measurement and x_k is the state to be estimated. The process noise ω_k and measurement noise e_k are assumed additive, zero mean and white noises and independent to each other. The initial state, with estimate $\bar{x}_0$, an approximation mean, and the associated covariance matrix P_0 which is positive definite, is assumed to be uncorrelated with the two noise sequences.

We assume the system (1.6) are subject to the following constraints:

$$x_k \in \Omega_x, \omega_k \in \Omega_\omega, e_k \in \Omega_e, \quad (1.7)$$

where Ω_x , Ω_ω and Ω_e are H-polytope which is defined in Section 2.1.3 containing the origin.

Given the most recent N measurements $y_{T-N+1}, y_{T-N+2}, \dots, y_T$ the constrained MHE with quadratic cost functions in [88, 89] is defined as follows,

$$\begin{aligned}
 J_T : \Phi_T^* = & \min_{\hat{x}_{T-N+1}, \{\hat{\omega}_k\}_{k=T-N+1}^T} \sum_{k=T-N+1}^T (\hat{e}_k' R^{-1} \hat{e}_k + \hat{\omega}_k' Q^{-1} \hat{\omega}_k) + \theta_{T-N+1}(\hat{x}_{T-N+1}), \\
 s.t. \quad & \hat{x}_{k+1} = f(\hat{x}_k, k) + \hat{\omega}_k, \\
 & y_k = h(\hat{x}_k, k) + \hat{e}_k, \\
 & \hat{x}_k \in \Omega_x, \hat{\omega}_k \in \Omega_\omega, \hat{e}_k \in \Omega_e.
 \end{aligned} \tag{1.8}$$

The optimal solution of the constrained MHE (1.8) is denoted as $\hat{x}_{T-N+1|T}^*$ and $\{\hat{\omega}_{k|T}^*\}_{k=T-N+1}^T$. Then the current state estimation $\hat{x}_{T|T}^*$ is the solution of the model (1.6) at time T subject to the initial condition $\hat{x}_{T-N+1|T}^*$ and disturbance sequence $\{\hat{\omega}_{k|T}^*\}_{k=T-N+1}^T$. The $\theta_{T-N+1}(\hat{x}_{T-N+1})$ is the arrival cost which denotes the contribution of the measurements outside of the window, i.e., $\{y_k\}_{k=0}^{k=T-N}$ on the estimation of state x_{T-N+1} and can be approximated by

$$\theta_k(z) = (z - \hat{x}_{k|k-1}^*)' P_k^{-1} (z - \hat{x}_{k|k-1}^*) + \Phi_{k-1}^*, \tag{1.9}$$

where P_k sequence is obtained by solving the matrix Riccati equation

$$P_k = Q + A_{k-1} P_{k-1} A_{k-1}' - A_{k-1} P_{k-1} C_k' (R + C_k P_{k-1} C_k')^{-1} C_k P_{k-1} A_{k-1}' \tag{1.10}$$

with initial condition P_0 and

$$A_k = \left. \frac{\partial f(x, k)}{\partial x} \right|_{x=\hat{x}_{k|k-1}^*}, \quad C_k = \left. \frac{\partial h(x, k)}{\partial x} \right|_{x=\hat{x}_{k|k-1}^*}$$

and

$$\hat{x}_{k|k-1}^* = f(\hat{x}_{k-1|k-1}^*, k-1).$$

In the next time instant when new measurement y_{T+1} occurs, the constrained optimization is repeated based on the measurements $y_{T-N+2}, y_{T-N+3}, \dots, y_{T+1}$ to get the estimation of state x_{T+1} .

In the next two subsections, as an illustration, we compare MHE with Kalman Filter and EKF when uncertainty exists, in particular, when we have poor prior knowledge about the process and measurement noises and strong ionosphere scintillation happens on GPS application.

1.2.1 Comparing MHE and Kalman Filter

Consider the following observable discrete time linear system:

$$x_{k+1} = \begin{bmatrix} 0.2695 & 0.4779 & 0.1127 & 0.1339 \\ 0.2688 & 0.4786 & 0.0753 & 0.1713 \\ 0.0053 & 0.0063 & 0.5713 & 0.0168 \\ 0.0035 & 0.0081 & 0.0094 & 0.5782 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} u_k + \omega_k,$$

$$y_k = \begin{bmatrix} 1 & 0.5 & 1 & 0.5 \end{bmatrix} x_k + e_k,$$

where y_k is the k^{th} measurement, x_k is the state to be estimated, ω_k is a zero mean truncated white Gaussian noise with the following variance and limit

$$E(\omega_k \omega_k') = 0.1 \times I_4, \quad \begin{bmatrix} -0.2 & -0.2 & -0.2 & -0.2 \end{bmatrix}' \leq w_k \leq \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 \end{bmatrix}'.$$

e_k is a zero mean white Gaussian noise with variance $E(e_k e_k') = 5$ and independent to ω_k . We assume the state is subject to the following constraint:

$$\begin{bmatrix} 2.2 & 2.4 & 0 & 0 \end{bmatrix}' \leq x_k \leq \begin{bmatrix} 4.3 & 4.2 & 2.7 & 2.3 \end{bmatrix}'.$$

Kalman Filter is an optimal estimation for linear system when Q and R are chosen

to match process and measurement noises. However, the variances of process and measurement noises are often not known exactly in practice. Q and R are often used as tuning parameters of the Kalman filter.

Figure 1.2 and 1.3 show that the Kalman Filter could produce estimates which violate the state constraints when Q and R are not chosen correctly while the constrained MHE gives state estimates that are within the constraints. The reason is that the prior information of constraints are taken into account in the constrained MHE while Kalman Filter does not.

Figure 1.2: Estimation performance with $Q = 10 \times I_4$ and $R = 5$.

Figure 1.3: Estimation performance with $Q = 0.1 \times I_4$ and $R = 0.1$.

Table 1.1: Performance Comparison of Kalman Filter and MHE

$Q = qI_4$ with q	R	Kalman Filter (J_p)	MHE with $N = 5$ (J_p)
0.1	5	13.9794	13.9794
10	5	166.6161	37.1588
10	1	194.8563	37.5834
100	1	203.5847	37.7831

Also shown in Table 1.1 we use the following criterion:

$$J_p = \sum_{k=1}^T (x_{1,k} - \hat{x}_{1,k})^2$$

where $\hat{x}_{1,k}$ is the estimate of $x_{1,k}$ of MHE or Kalman Filter to compare the estimation performance of Kalman Filter and MHE when Q and R change. The variances of process and measurement noises are $E(\omega_k \omega_k') = 0.1I_4$ and $E(e_k e_k') = 5$ (row 2 in Table 1.1). When this is the case, the performance of Kalman Filter and constrained MHE is almost the same. The performance of Kalman Filter deteriorates significantly when the values of q and R are inaccurate (rows 3 to 5 in Table 1.1). The constrained MHE, however, is less sensitive to the values of q or R . Thus, constrained MHE is less affected by uncertain information, in particular by the inaccuracy of process and measurement noise variances..

1.2.2 Comparing MHE and EKF

In order to simplify the solving of the constrained optimization problem in MHE (1.8) we use the following linearization version of (1.8):

$$\begin{aligned}
J'_T : \Phi_T^* &= \min_{\hat{x}_{T-N+1}, \{\hat{\omega}_k\}_{k=T-N+1}^T} \sum_{k=T-N+1}^T (\hat{e}_k' R^{-1} \hat{e}_k + \hat{\omega}_k' Q^{-1} \hat{\omega}_k) + \theta_{T-N+1}(\hat{x}_{T-N+1}), \\
s.t. \quad &\hat{x}_{k+1} = f(\hat{x}_{k|k-1}^*, k) + A_k(\hat{x}_k - \hat{x}_{k|k-1}^*) + \hat{\omega}_k, \\
&y_k = h(\hat{x}_{k|k-1}^*, k) + C_k(\hat{x}_k - \hat{x}_{k|k-1}^*) + \hat{e}_k,
\end{aligned}$$

$$\hat{x}_k \in \Omega_x, \hat{\omega}_k \in \Omega_\omega, \hat{e}_k \in \Omega_e. \quad (1.11)$$

where

$$\begin{aligned} \theta_{T-N+1}(\hat{x}_{T-N+1}) &= (\hat{x}_{T-N+1} - \hat{x}_{T-N+1|T-N}^*)' P_{T-N+1}^{-1} (\hat{x}_{T-N+1} - \hat{x}_{T-N+1|T-N}^*), \\ A_k &= \left. \frac{\partial f(x, k)}{\partial x} \right|_{x=\hat{x}_{k|k-1}^*}, \quad C_k = \left. \frac{\partial h(x, k)}{\partial x} \right|_{x=\hat{x}_{k|k-1}^*}. \end{aligned}$$

The constrained optimization problem (1.11) is a constrained Quadratic Programming (QP) problem which can be solved by MATLAB command “quadprog()”.

Radar application

Consider the following radar tracking application

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + \omega_k, \\ y_k &= \begin{bmatrix} \sqrt{x_{1,k}^2 + x_{3,k}^2} \\ \arctan \frac{x_{3,k}}{x_{1,k}} \end{bmatrix} + e_k, \end{aligned}$$

where the state is the positions and velocities of the object in North-South (N-S) and East-West (E-W) directions. y_k is the k^{th} measurement. ω_k and e_k are zero mean white Gaussian noises with variances

$$E(\omega_k \omega_k') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.01 \end{bmatrix}, \quad E(e_k e_k') = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$$

and independent to each other. We assume the velocities in N-S and E-W directions are subjected to the following constraints:

$$140\text{m/s} \leq x_{2,k}, \quad x_{4,k} \leq 210\text{m/s}.$$

Suppose the initial position of the object is in the origin and the initial velocities in N-S and E-W directions are both 200m/s, i.e., $x_0 = [0 \ 200 \ 0 \ 200]'$.

Figure 1.4 shows that the performance of EKF and constrained MHE are almost the same when Q and R are chosen correctly. Constrained MHE is able to keep the state estimates within the constraints while the EKF is unable to do so shown in Figure 1.5 when the values of Q and R are not chosen correctly. The inaccurate values of Q and R used in Figure 1.5 are

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R = \begin{bmatrix} 0.09 & 0 \\ 0 & 0.1 \end{bmatrix}.$$

Figure 1.4: Estimation of Velocities using EKF and constrained MHE when Q and R are chosen correctly.

Figure 1.5: Estimation of Velocities using EKF and constrained MHE when Q and R are inaccurate.

GPS application

Vector tracking is an approach to simultaneously process Global Navigation Satellite System (GNSS) signal tracking and position, velocity, time (PVT) estimation, commonly based on an extend Kalman filter (EKF). In vector tracking, bad measurement data (or fault) in one channel will affect the other channels, and in the worst case, could lead to loss of lock in all channels. With the capability to take constraints into account in the estimation, MHE can constrain the effect of each channel within a priori defined range, thus mitigate the effect from the faulty channels [107]. Figure 1.6 shows the doppler frequency estimations of MHE-Vector tracking and EKF-Vector tracking for 6 satellites under strong ionosphere scintillation. From Figure 1.6 we can see that EKF-Vector fails to track the correct carrier frequency, and a false frequency lock presents at around 90 s. The false lock is due to the amplitude loss caused by ionosphere scintillation. As comparison, the estimates from MHE-Vector tracking are barely affected by the scintillation. Therefore, MHE-Vector tracking gives a better performance than EKF under ionosphere scintillation.

From the above three simulations we conclude that, with constraints, MHE can

Figure 1.6: Doppler Frequency Estimation.

produce better estimation comparing to Kalman Filter and EKF when poor knowledge of process and measurement noises is assumed or strong ionosphere scintillation happens in GPS application. As constraints are taken into account, a constrained optimization problem in MHE must be solved at each time instant. The feasibility of MHE plays a key role in the implementation of MHE. By the symmetry exposed in [76], the constrained MHE can be translated to an equivalent MPC with tracking error bound. In this thesis, by leveraging the symmetry exposed in [76], the feasible set of the equivalent MPC then becomes the feasible set of the MHE and the counter part of the trajectory advisor for the MPC is a measurement checker for the MHE. The measurement checker could be used to determine the feasibility of the MHE before the actual constrained optimization is carried out.

1.2.3 An influence function approach to handle non-Gaussian noise assumption

Another way to tackle non Gaussian noise in estimation is to use influence function [47]. The influence function can tell us the robustness of different existing estimators and be used to design new robust estimators. In addition, it is well-known that t -distribution [62] has the property of thick tail to better model the occur of outliers and reduces to Gaussian distribution when its shape parameter of distribution tends to infinity. Thus t -distribution has the flexibility to characterize noise with Gaussian or non-Gaussian statistical properties. In the literature the heavy tail property of t -distribution has been used to improve the performance of estimators in [1, 2, 79, 95]. These papers are based on Bayesian method and the heavy tail property of t -distribution was used to down weigh the outliers. In the application of indoor environment localization, the measurement noise is not Gaussian noise as shown in Figure 1.7 where the measurement noise is better fitted using a t -distribution instead of a Gaussian distribution. In this thesis, our method

Figure 1.7: The histogram of measurement noise of indoor environment localization.

differs from these estimators from [1, 2, 79, 95] in that the probability density function (pdf) of t -distribution was used directly in the likelihood function of Maximum Likelihood Estimation (MLE) by using the most recent N measurements. MLE for non-Gaussian pdf is a nonlinear problem. The proposed influence function approach is that we used the Influence Function (IF) to obtain an approximate solution to the original MLE problem. More specifically, we obtained an approximate solution to the MLE for Autoregressive-Moving-Average with eXogenous (ARMAX) process with t -distribution noise model. The approximation solution results a computationally efficient Kalman filter like recursive algorithm parameterized by the shape parameter of t -distribution. The recursive solution reduces to the traditional Moving Window Least-Squares Estimation when the noise is Gaussian. The moving horizon N can be used as tuning parameter for the proposed estimator. The formula for the variance of the estimate is derived. This formula shows explicitly how the variance of estimate is affected by the number of measurements and noise variance. In addition, a formula which predicts how the effect of outlier on the estimate is derived. Other approaches for handling non-Gaussian noise include particle filter [8] which is based on Monte Carlo method, but the heavy computation load of particle filter limits its applications.

1.3 Contributions

The contributions of this thesis are:

- We showed relationship between a trajectory advisor for MPC with tracking error bound and a measurement checker for constrained MHE for constrained linear systems. Through symmetry between the MPC with tracking error bound and constrained MHE, the counter part of the trajectory advisor is the measurement checker.

-
- The trajectory advisor can supply a new feasible reference signal to recover the feasibility of the MPC problem if the original reference signal makes the MPC problem infeasible. The illustrated examples show that by including tracking error bound, MPC gives faster transient than without tracking error bound. Incremental input model and suitable modification of the cost function are needed to achieve zero offset in the presence of constant disturbance. The measurement checker could be used to determine the admissibility of the measurements before the constrained MHE is carried out. Some subtle differences between MPC and MHE are also highlighted.
 - Proposed a MHE for ARMAX process with t -distribution noise model. An influence function approach is employed to obtain an approximate solution for the proposed MHE. The approximation solution results a computationally efficient Kalman filter like recursive algorithm parameterized by the shape parameter of t -distribution. When the noise is Gaussian the proposed MHE reduces to standard MHE.
 - The statistical properties of the proposed MHE are analyzed. Formulas of estimation variance and the effect of the outlier on the estimate are derived and verified by simulations. Simulation results show that the proposed estimator gives smaller estimation variance and is less affected by outlier than Moving Window Least-Squares Estimation , and more robust to modeling error than Kalman Filter.
 - The proposed MHE is applied to indoor environment localization and liquid level estimation applications. The experimental results show that the proposed MHE is less affected by outlier and gives smaller estimation variance than Moving Window Least-Squares Estimation.

The remaining of the thesis is organized as follows. In Chapter 2 we give the necessary background. The effect of S/H time on the size of initial feasible set in MPC design is investigated and observed through illustrative examples in Chapter 3. The relationship between the initial feasible set and S/H time was used to select an appropriate S/H time for MPC while taking into account closed-loop performance and computational complexity of MPC, for the case when trade-off between implementation cost and performance benefits is required. In Chapter 4 a trajectory advisor for MPC with tracking error bound and a measurement checker for constrained MHE are proposed. MHE for ARMAX process with t -Distribution noise by using influence function is proposed in Chapter 5. Chapter 6 concludes this thesis and gives some recommendations for future work.

Chapter 2

Background

2.1 Model Predictive Control

In the past several decades, MPC has been developed extensively. It has been widely applied in industry process control. Some MPC software products have been developed, such as the Matlab MPC toolbox and Multi-Parametric toolbox [52]. Also several well known books [4, 24, 75, 92, 106] about MPC were published.

2.1.1 Recursive Feasibility of MPC

The invariant set theory [19] is essential to guarantee the recursive feasibility of MPC which was discussed early in [12, 13, 18, 44, 103]. Given a dynamic system a subset of the state space is said to be positively invariant if it has the property that, if it contains the system state at some time, then it always contains the system state [19] in the future time. In order to guarantee recursive feasibility of MPC, a terminal invariant set is usually designed in MPC. The initial feasible set then is defined as a set of states for which there exists an admissible control sequence which will drive the state into the terminal invariant set in N steps or less while keeping the evolution of the state inside the state constraints.

Standard approaches to guarantee the recursive feasibility of MPC use the dual-mode prediction paradigm [78, 97] where the terminal invariant set is designed. The initial feasible set is calculated backward from the terminal invariant set. The terminal invariant set, often associated with a terminal cost, is used in the design of MPC. Such technology not only retains the known advantages of an infinite horizon, such as stability and robustness, but also guarantees the recursive feasibility of the constrained optimization problem.

The conditions which ensure closed-loop asymptotic (exponential) stability and recursive feasibility of the standard MPC (1.3) [78] are

1. : $\mathbb{X}_f \subseteq \mathbb{X}$, $0 \in \mathbb{X}_f$ (state constraint satisfied in $\mathbb{X}_f$).
2. : $K_f(x) \in \mathbb{U}$, $\forall x \in \mathbb{X}_f$ (control constraint satisfied in $\mathbb{X}_f$).
3. : $f(x, K_f(x)) \in \mathbb{X}_f$, $\forall x \in \mathbb{X}_f$ ($\mathbb{X}_f$ is positively invariant set under $K_f(\cdot)$).
4. : $\left[\dot{F}_L + \ell \right] (x, K_f(x)) \leq 0$, $\forall x \in \mathbb{X}_f$ ($F_L(\cdot)$ is a local Lyapunov function).

where $K_f(x)$ is a stable controller for the system (1.1) in a neighbourhood of origin. $\dot{F}_L(\cdot)$ is denoted as the first-order derivative of $F_L(\cdot)$.

Here the idea of ensuring the stability of constrained MPC is that once the state is steered into the terminal invariant set $\mathbb{X}_f$ by the optimal control one can switch to the Lyapunov stable controller $K_f(\cdot)$; the so-called dual-model MPC [97, 98]. MPC with constraints implicitly assumes that the state of the plant lies within the initial feasible set which can be calculated backward from the terminal invariant set $\mathbb{X}_f$.

For linear system ($f(x_k, u_k) = Ax_k + Bu_k$) where (A, B) is stabilizable, $\ell(x_k, u_k)$ usually is $\|x_k\|_{Q_0}^2 + \|u_k\|_{R_0}^2$ where $Q_0 \geq 0$, $R_0 > 0$, $F(x)$ and $K_f(x)$ are often chosen as the optimal value function and optimal controller for the infinite unconstrained LQR (so $F(\cdot)$ and $K_f(\cdot)$ are easily computed), respectively. The terminal constraint set $\mathbb{X}_f$ is chosen as the maximal output admissible set [37] of the closed-loop system $x_{k+1} = (A + BK_f)x_k$.

DEFINITION 2.1 TERMINAL INVARIANT SET $\mathbb{X}_f$ [78]. *The terminal invariant set for linear system $Ax_k + Bu_k$ under infinite unconstrained LQR controller K_f is defined as*

$$\mathbb{X}_f \triangleq \{x \in \mathbb{X} : K_f x \in \mathbb{U} \text{ and } (A + BK_f)x \in \mathbb{X}_f\},$$

where

$$K_f = -(R_0 + B'PB)^{-1}B'PA \quad (2.1)$$

and

$$P = A'PA - A'PB(B'PB + R_0)^{-1}B'PA + Q_0. \quad (2.2)$$

DEFINITION 2.2 INITIAL FEASIBLE SET [60]. *The set $\mathbb{X}_0(N, \mathbb{X}_f)$ is the initial feasible set contained in $\mathbb{X}$ for the system $x_{k+1} = Ax_k + Bu_k$ if and only if $\mathbb{X}_0(N, \mathbb{X}_f)$ contains all the states in $\mathbb{X}$ for which there exists an admissible control law which will drive the state of the system into $\mathbb{X}_f$ in N steps or less, while keeping the evolution of the state inside $\mathbb{X}$, i.e.*

$$\mathbb{X}_0(N, \mathbb{X}_f) \triangleq \{x_0 \in \mathbb{X} : \exists \{u_k \in \mathbb{U}\}_0^{N-1}, M \leq N \quad (2.3)$$

$$\text{such that } \{x_k \in \mathbb{X}\}_0^{M-1} \text{ and } \{x_k \in \mathbb{X}_f \subseteq \mathbb{X}\}_M^N\}.$$

THEOREM 2.1 *The initial feasible set $\mathbb{X}_0(N, \mathbb{X}_f)$ is a positively invariant set [78].*

Proof. See Appendix 1. □

Thus for linear system the MPC (1.3) often reduces to

$$\begin{aligned} \min_{\{u_k\}_{k=0}^{N-1}} \quad & x_N' P x_N + \sum_{k=0}^{N-1} \|u_k\|_{R_0}^2 + \|x_k\|_{Q_0}^2, \\ \text{s.t.} \quad & x_{k+1} = Ax_k + Bu_k, \quad (k = 0, 1, \dots, N-1) \\ & x_k \in \mathbb{X}, \quad (k = 0, 1, \dots, N) \end{aligned}$$

$$\begin{aligned}
u_k &\in \mathbb{U}, \quad (k = 0, 1, \dots, N-1) \\
x_N &\in \mathbb{X}_f \subseteq \mathbb{X}.
\end{aligned} \tag{2.4}$$

DEFINITION 2.3 RECURSIVE FEASIBILITY OF MPC. *MPC (2.4) is recursively feasible if and only if the initial state x_0 is inside the initial feasible set $\mathbb{X}_0(N, \mathbb{X}_f)$.*

The approximated computation methods of the (robust) initial feasible set are investigated in [14, 57, 60, 63, 96]. This thesis does not discuss the computational complexity of the initial feasible set.

2.1.2 Feasibility of MPC tracking problem

Besides regulation, MPC is often required to drive a system output to track a time varying reference, such as piecewise constant references, or a target set (zone) while ensuring that the state and control constraints, and even bound tracking error are satisfied all the times. Changes in reference signals may cause MPC to have no solution. Similar to the recursive feasibility of MPC regulation problem discussed above, the invariant set theory also provides a very useful tool to guarantee the feasibility of MPC tracking problem. Based on the theory of invariant set some works have been done on the feasibility of MPC tracking problem.

MPC for tracking piecewise constant references has been discussed in [16, 26, 33, 70, 71, 84, 92]. A novel method of MPC tracking problem was proposed in [5, 35, 71] where the controller features that an artificial steady state and input are considered as decision variables. An additional offset cost, i.e., the cost function penalizes the deviation between the artificial steady state and the desired one, is introduced in the cost function. The stability of the controller is guaranteed by adding a tracking error penalty term and a terminal invariant constraint for both the terminal state and the artificial steady state and input to the constrained optimization. The controller ensures feasibility for all admissible references and drives the system to

the admissible reference as close as possible. [36] solved the loss of the optimality property in [71] due to the addition of the artificial steady-state together with the proposed cost function by replacing the offset cost with an infinite norm cost function, and the robust version of [71] which can handle additive disturbances by using tube-based MPC is investigated in [70].

For tracking non piecewise constant time varying reference the MPC schemes have been discussed in [15, 31, 82]. In [15] the maximal invariant set is used to construct an invariant set for feasible reference signal and two approximate computation methods are proposed to calculate the maximal invariant set. In [31] a MPC scheme for tracking a bound random reference is proposed and the set to which the tracking error converges and the associated region of convergence are characterized. In [82] the feasibility of MPC tracking problem is guaranteed by employing augmented model and terminal conditions. In order to recover the feasibility a reference governor is designed to provide an admissible reference whenever it is necessary. In addition the explicit formulations of predictive law and those of the reference governor mechanism are mixed together to give a compact explicit MPC law with guarantees of feasibility for reference tracking. In this thesis a trajectory advisor for a standard MPC tracking problem with tracking error bound is designed.

Another way to address the feasibility of constrained tracking problem is reference governor which was discussed in [13, 38, 39, 41, 61, 101, 102] and the references therein. The strategy of reference governor is to separate the issue of constraint satisfaction from the issue of designing a stable closed-loop system. The basic idea of the reference governor is to add an auxiliary nonlinear device between the reference command input and the closed-loop control system as shown in 2.1 [38].

Figure 2.1: Reference Governor.

A wide range of different controller synthesis methods can be applied to the plant to yield a closed-loop system that performs satisfactorily in absence of constraints. Reference governor selects an admissible reference y^{r*} online whenever it is necessary based on the current state x , reference signal y^r and the constraints on state and input to avoid the constraint violation.

The idea that unifies the process economic optimization and process control by including economic objective directly in an MPC scheme is called Economic MPC (EMPC) [29, 93]. In the process industries, the traditional paradigm for achieving economic objectives is to partition the information management, decision making and system control into two layers. The first layer, often referred as real-time optimization (RTO), performs a steady-state economic optimization of the plant's variables and sends the results of its optimization as a setpoint to the second layer which usually referred as the advanced process control, such as, MPC. The task of the second layer is to guide the plant's transient state to the setpoint, and once arrived, to reject dynamic disturbances that enter the system. Alternative to the hierarchical decomposition of economic optimization and process control, EMPC replaces the tracking stage cost function directly with the economic stage cost function used in the RTO layer (or elsewhere). Thus EMPC directly and dynamically optimizes the economic operating cost of the process, doing so without reference to any

steady state. In recent years EMPC has received much interest [6, 7, 28, 51, 91], and EMPC has also been applied to building energy systems [74] and power management in SmartGrid [54].

2.1.3 H-Polytope and Its Operations

Constraints arise in virtually a lot of control and estimation problems. For example, consider the case of automobile control. It is well known that there exist maximum and minimum available throttle displacement and braking capacity, that is, the input u to the system is constrained. Another example is chemical process control where valves have a maximum displacement (when fully open) and minimum displacement (when fully closed). Also, it is usual to impose limits on allowable temperature and pressure. In the estimation of concentration in a distillation column, it is typically known that the liquid levels in the trays must lie between given lower and upper limits. All these constraints can be expressed in the form of H-polytope. The constraints considered in this thesis are assumed to be H-polytopes which consist of a set of inequalities. The definition of H-polytope and its operations are defined as follows.

DEFINITION 2.4 H-POLYTOPE [111]. *A H-polytope is a bounded H-polyhedron. A H-polyhedron is an intersection of closed halfspaces which is a set $P \subseteq \mathbb{R}^d$ presented in the form*

$$P = \{x \in \mathbb{R}^d : Hx \leq f\}, \text{ for some } H \in \mathbb{R}^{m \times d}, f \in \mathbb{R}^m.$$

DEFINITION 2.5 ORTHOGONAL PROJECTIONS OF H-POLYTOPE [96]. *The orthogonal projections of a H-polytope $P \subset \mathbb{R}^{n+d}$ onto its first n coordinates $\mathbb{R}^n$ and last*

d coordinates $\mathbb{R}^d$ are defined as

$$\Pi_n^1(P) = \left\{ x \in \mathbb{R}^n : \exists z \in \mathbb{R}^d, \begin{bmatrix} x \\ z \end{bmatrix} \in P \right\}$$

and

$$\Pi_d^2(P) = \left\{ z \in \mathbb{R}^d : \exists x \in \mathbb{R}^n, \begin{bmatrix} x \\ z \end{bmatrix} \in P \right\},$$

respectively.

DEFINITION 2.6 AFFINE TRANSFORMATION OF POLYTOPE [52]. $P = \text{range}(Q, A, f) = \{Ax + f \in \mathbb{R}^n | x \in Q\}$ where $Q \subset \mathbb{R}^m$ is a H -polytope, $A \in \mathbb{R}^{n \times m}$ and $f \in \mathbb{R}^n$.

DEFINITION 2.7 MINKOWSKI SUM [60]. Given two H -polytopes $\Omega \subset \mathbb{R}^n$ and $\Phi \subset \mathbb{R}^n$, the Minkowski sum of Ω and Φ is defined as

$$\Omega \oplus \Phi \triangleq \{x \in \mathbb{R}^n | \exists \omega \in \Omega, \phi \in \Phi : x = \omega + \phi\}.$$

2.2 Moving Horizon Estimation

In literature constrained estimation has been discussed a lot, such as in [3, 73, 88, 89, 94]. Constrained MHE for linear and nonlinear systems, and the conditions of nominal asymptotically stability are investigated in [88, 89]. The advantages of the constrained MHE over EKF for nonlinear system are shown in [49] for chemical reaction systems and two conditions under which the EKF fails are proposed. The research of [110], [65], [50] and [27] focused on overcoming the difficulty to solve a constrained optimization problem in constrained MHE. In practice, the constrained MHE has been applied to some applications ([32], [86] and [85]). In this thesis, we focus on the constrained MHE investigated in [88, 89]. The definition of the constrained MHE is given by (1.6) in Chapter 1.

2.3 Symmetry between control and estimation

The Lagrange duality between Linear Quadratic Regulator (LQR) and Kalman filter for unconstrained system is well understood [58, 99], and the duality can be summarized as the observations that gain and Riccati equation of Kalman filter are the same as those of the LQR with exception that initial conditions are given for Riccati equation for filter solution, while final conditions are given for Riccati equation for the regulator solution. For constraint case, the Lagrangian duality between constrained control and constrained estimation is investigated [42]. The constrained estimation problem is converted to a nonlinear unconstrained control problem by the Lagrangian duality and the duality is shown through projected inputs of the constrained estimation and projected outputs of the nonlinear control problem.

On another hand, [100] shows that the continuous time Kalman Filter can be translated to Linear Quadratic Tracking (LQT) problem in reversing time. This process can be obtained by first reversing time, so that the cost is imposed on final state, then minimizing over all possible solutions to the LQT problem, over all possible initial states. Based on this relationship a similar symmetry between the formulation of discrete time constrained state estimation and constrained reference tracking problem is exposed in [76]. The symmetry relationship allows one to interpret a constrained state estimation problem as a constrained reference tracking problem, and vice versa. This thesis focuses on the symmetry exposed in [76].

In addition, in Appendix 2, an alternative derivation of Kalman filter from LQT by symmetry for unconstrained discrete time system is given.

2.4 Limitation of Kalman filter

Kalman filter is the most popular method for state estimation. By Bayesian law Kalman filter is the best mean square error estimator for linear system and Gaussian noise assumption. Specifically prior distribution and likelihood distribution both are Gaussian, and the posterior distribution turns out to be also Gaussian which is a compromise between the prior and the likelihood distributions. As a result simplicity is the feature of Kalman filter. However, Kalman filter is not robust to outliers because the discrepancy between the observation and its prior predicted value is unbounded and increasing linearly. In the case of outliers the heavy tail distribution should be considered, such as Student t-distribution [62] which is symmetric and behaves much like Gaussian in the central area but has heavier tails. However, updating the posterior distributions with non-Gaussian noise assumption is usually analytically intractable.

In recent years, from the distribution-theoretic point of view, some methods were proposed to approximate the updates of posterior distributions. In [1, 2] the posterior distribution is approximated by Gaussian noise in the sense of minimizing a so-called Kullback-Leibler (KL) divergence which is a natural distance metric between probability distributions. In [79] the posterior distribution is assumed to be a mixture of Student t-distribution. In [95] the conditional joint distributions between state and process noise, state and measurement noise were approximated by Student t-distribution to preserve the heavy tails, etc. The posterior distribution then is obtained by linear transformation. The above robust methods make use of the heavy tail of the assumed non-Gaussian noise (such as, Student t-distributed noise) to down weigh the outliers in the estimator. An optimization approach using ℓ_1 regularization was proposed to handle an additional measurement noise term which is sparse in [77]. The additional measurement noise is used to model the outliers. In [80] the Student t-distribution was used to discretize between the Gaussian noise

and outliers.

2.5 Influence Function

Bayesian estimation is based on theories of classical parametric statistics and an optimal procedure under exact parametric models. In order to explore the performance of estimation when the models are only approximately valid (or, say the performance of the estimation in the neighborhood of the assumed parametric model) Hampel [46, 47] proposed the concept of influence function based on the pioneer work done by Huber in [55] where a rather full kind of neighborhood of a strict parametric model was considered first time. The theory of influence function provides a framework for judging and comparing the existing statistical procedures under the aspect of robustness, and finally develops and suggests new robust procedures (such as, Redescending M-estimator [48]).

Suppose n one-dimensional observations $Y_1, \dots, Y_n$ are independent and identically distributed (i.i.d). The observations belong to some sample space on which a family of cumulative probability distributions F_θ is defined, where the unknown parameter θ belongs to some parameter space. Consider real value statistics (estimators)

$$T_n = T(Y_1, \dots, Y_n) = T_n(G_n)$$

where G_n is the empirical distribution of the observations and given by $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{y_i}$ where $\mathbf{1}_{y_i}$ is an unit step function at the point y_i . We consider estimators which are functionals, i.e., $T_n(G_n) = T(G_n)$ for all n and G_n , or can be asymptotically be replaced by functionals. Here we assume that the functionals under study are Fisher consistent :

$$T(F_\theta) = \theta \text{ for all } \theta$$

which means that at the model, i.e., the observations are according to the true distribution F_θ in domain (T) , the estimators $\{T_n; n \geq 1\}$ asymptotically measures the right quantity. Assume that T is Gâteaux differentiable at distribution F in domain (T) , the influence function is given by the following definition.

DEFINITION 2.8 INFLUENCE FUNCTION [47]. *The Influence Function (IF) of T at F is given by*

$$\text{IF}(y; T, F) = \lim_{h \rightarrow 0} \frac{T((1-h)F + h\Delta_y) - T(F)}{h} \quad (2.5)$$

where $h \in [0, 1]$ which is the mass of contamination to F , y is on the sample space where this limit exists and Δ_y is probability measure which puts mass 1 at the point y .

The importance of the IF lies in its heuristic interpretation: it describes the effect of an infinitesimal contamination at the point y on the estimate, standardized by the mass of the contamination. If we say the point y is an outlier the IF indeed describes the effect of an outlier on the estimation. Thus from the IF point of view, the IF of a robust estimator should be bound and/or descending, or increasing slowly for large magnitude of noise (also remarked in [9] the robust estimator may be obtained by mapping the estimation error such that the nonlinear map function is linear for small estimation error but increases more slowly than linear for large estimation error). Based on this principle Hampel suggested a Redescending M-estimator whose IF reduces to zero again for gross outliers which means the estimator completely rejects the gross outliers.

The important relationship between the IF and asymptotic variance is that when the observations Y_i are i.i.d. according to F the distribution of $\sqrt{n}(T_n - T(F))$ tends to normal distribution $N(0, V(T, F))$ as n goes to infinity, where the asymptotic

variance equals

$$V(T, F) = \int \text{IF}(y; T, F)^2 dF(y) \quad (2.6)$$

Also some robustness measures can be derived from IF. The gross-error sensitivity of T is defined as the supremum of the absolute value of IF and regarded as an upper bound on the (standardized) asymptotic bias of the estimator. The local-shift sensitivity of T is defined as the supremum of the slope of the IF and measures the worst effect of "wiggling" around y . The rejection point of T is defined as the infimum of the absolute value of y from which the IF is always zero. Thus the rejection point is the smallest gross outliers which are rejected by the estimator completely.

2.6 ARMAX process

The ARMAX process is popularly used in the statistical analysis of time series [10, 20]. The Kalman filter of ARMAX was discussed in Example 11.6 in [10]. The single-input single-output ARMAX process is given by:

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k \quad (2.7)$$

where

$$\begin{aligned} A(q^{-1}) &= 1 + a_1q^{-1} + \cdots + a_nq^{-n} \\ B(q^{-1}) &= b_1q^{-1} + b_2q^{-2} + \cdots + b_{n_B}q^{-n_B} \\ C(q^{-1}) &= 1 + c_1q^{-1} + \cdots + c_nq^{-n} \end{aligned}$$

$k = 1, \dots, T$ is the sampling instance, $n_B \leq n$ and q^{-1} is the backward shift operator, i.e., $q^{-1}y_k = y_{k-1}$. a_i , b_i and c_i are the parameters of polynomials $A(q^{-1})$, $B(q^{-1})$

and $C(q^{-1})$, respectively. The polynomial $C(q^{-1})$ has all its zeros inside the unit disc and may be multiplied by an arbitrary power of q as this does not change the correlation structure of $C(q^{-1})$ [90]. This is used to normalized C so that $\deg C = \deg A = n$. The input and output are given by u_k and y_k respectively. And e_k is zero mean white measurement noise.

Next based on the background given in this chapter we introduce our work in the following chapters.

Chapter 3

The Effect of Sample/Hold Time on Initial Feasible Set in Model Predictive Control Design

S/H time plays an important role in the implementation of MPC. The purpose of this chapter is to investigate and observe how the S/H time effects the size of initial feasible set in MPC design. Many factors influence S/H time choice, e.g. the process dynamics, the computational power of the implementation platform, etc. An illustrated example is given to show how to select S/H time in MPC designs, especially from the perspective of trade-off between their on-line computational complexity, initial feasible set, and closed-loop performance.

3.1 Introduction

Model Predictive Control (MPC) solves online, a finite dimension optimal problem at each sampling instant given the current state of the plant; the optimization yields an optimal control sequence and the first control in this sequence is applied to the plant, in order to generate feedback to possible disturbance [75]. An important

advantage of MPC is its ability to handle constraints on controls and states.

Originally developed to meet the specialized control needs of power plants and petroleum refineries, MPC technology can now be found in a wide variety of application areas including chemicals, food processing, automotive and aerospace applications [87]. Terminal cost and constraint set are used prevalently in the cost function of MPC [78] primarily to retain the known advantages of an infinite horizon, such as stability and robustness, and also guarantee the recursive feasibility of the optimization problem. Here the idea of ensuring the stability of constrained MPC is that once the state is steered into the terminal constraint set by the optimal control one can switch to a Lyapunov stable controller; the so-called dual-model MPC [97]. MPC with constraints implicitly assumes that the state of the plant lies within the so-called initial feasible set which is a set of initial state for which there exists an admissible control sequence which satisfies the state and control constraints and ensures the terminal constraint set reached in N steps [60]. So the initial feasible set is largest subset of state space in which there exists a control sequence satisfying the control and state constraints and is the domain where the optimal control function is defined. The initial feasible set is related to the prediction horizon and the terminal set. Longer prediction horizon or bigger terminal set usually results in larger initial feasible set.

The initial feasible set plays a key role in the recursive feasibility of MPC. In the design of MPC recursive feasibility is very important to prevent the state from being driven into a region where the constrained optimization problem has no solution (i.e. there doesn't exist a sequence of control inputs satisfying all constraints). Thus the initial feasible set is often required as a positively invariant set to ensure the recursive feasibility of MPC. Given a dynamic system a subset of the state space is said to positively invariant if it has the property that, if it contains the system state at some time, then it will contain the system state also in the future [19]. In

recent years approximated computation methods of the (robust) initial feasible set are investigated in [14, 60, 63, 96]. From application perspective, the initial feasible set can be useful in the design of MPC since it gives information about the region of the state space where the MPC can be applied. In this chapter we take initial feasible set into account in the selection of Sample/Hold (S/H) time of MPC design.

Given a continuous time cost function, it is well known that, under the Zero-Order-Hold (ZOH) discretization, the discrete-time cost function will include a cross term of state and input, and the weights of state and input are functions of the S/H time. In conventional discrete-time design of MPC, the weights are usually not adjusted when h changes, and the cross term is usually omitted, introducing an inherent quadrature error [83]. In this chapter we formulate MPC based on continuous time cost function to account for the discretization effect when S/H time changes.

This chapter is organized as follows. In Section 3.2, we formulate the discrete-time MPC problem based on continuous time cost function and the parameters (weights) of the discrete-time MPC problem are functions of S/H time. In Section 3.3 we show how the initial feasible set changes when the S/H time changes. In Section 3.4 an illustrated example is given to show how to select S/H time from perspective of trade-off between closed-loop performance, initial feasible set and computational complexity for constrained MPC. Section 3.5 concludes this chapter.

3.2 MPC Design Based on Continuous Time Cost Function

Consider a linear continuous-time state space model of a plant

$$\dot{x}_t = A_c x_t + B_c u_t, \quad (3.1)$$

where $u_t \in \mathbb{R}^m$ and $x_t \in \mathbb{R}^n$ represent the inputs and internal states, respectively and (A_c, B_c) is a stabilizable pair.

Given (3.1) the continuous time cost function is defined as:

$$\begin{aligned} J &= x_T' P_c x_T + \int_0^T (x_t' Q_c x_t + u_t' R_c u_t) dt, \\ \text{s.t. } u_t &= u_{ih}, \quad \forall t \in [kh, (k+1)h), \end{aligned} \quad (3.2)$$

where T is the prediction horizon measured in time, h is the S/H time and k is the sampling instant, Q_c is positive semi-definite, P_c and R_c are positive definite, $(\sqrt{Q_c}, A_c)$ is detectable. Current state x_t is assumed to be known at time t . One possible choice of P_c is the stabilizing solution to the following Riccati equation

$$P_c A_c + A_c' P_c - P_c B_c R_c^{-1} B_c' P_c + Q = 0. \quad (3.3)$$

3.2.1 Discretization of the continuous-time state-space model

For simplicity, let k and $k+1$ denote kh and $(k+1)h$, respectively. Then the system in (3.1) can be expressed as a discrete time system

$$x_{k+1} = A x_k + B u_k, \quad (3.4)$$

where

$$A = e^{A_c h}, \quad B = \int_0^h e^{A_c \tau} d\tau B_c. \quad (3.5)$$

If h is non-pathological (i.e. A_c does not have two eigenvalues with equal real parts and imaginary part that differ by an integral multiple of $2\pi/h$ [25], Theorem 3.2.1), then (A_c, B_c) controllable implies (A, B) reachable.

3.2.2 Discretization of the cost function

In this chapter, we only consider the S/H time h which is divisible by the continuous time prediction horizon T . The general case can be found in [11]. The cost function in (3.2) is expressed in continuous time. Integrating the continuous time cost function over intervals of length h when the control input is held constant gives

$$J = F(x_N) + \sum_{k=0}^{N-1} \begin{bmatrix} x_k \\ u_k \end{bmatrix}' Q_h \begin{bmatrix} x_k \\ u_k \end{bmatrix}, \quad (3.6)$$

where

$$F(x_N) = x_N' P_c x_N, \quad (3.7)$$

$$Q_h = \begin{bmatrix} Q_0 & S_0 \\ S_0' & R_0 \end{bmatrix}, \quad (3.8)$$

$$\begin{aligned} Q_0 &= \int_0^h (e^{A_c t})' Q_c e^{A_c t} dt, \\ S_0 &= \int_0^h ((e^{A_c t})' Q_c \int_0^t e^{A_c \tau} d\tau B_c) dt, \\ R_0 &= \int_0^h ((\int_0^t e^{A_c \tau} d\tau B_c)' Q_c \int_0^t e^{A_c \tau} d\tau B_c + R_c) dt. \end{aligned} \quad (3.9)$$

The discrete time prediction horizon step N is defined as

$$N := \frac{T}{h}. \quad (3.10)$$

With the assumption that u_t is constant over the interval $t \in [ih, (i+1)h)$ and state x_k is sampled at time instant $t = ih$ there is no approximation error between continuous time cost function and the sampled cost function.

As a summary, the ZOH discrete-time equivalent of MPC, given continuous-time

cost function (3.2) and continuous-time model (3.1) is:

$$\begin{aligned}
\min_{\{u_k\}_{k=0}^{N-1}} \quad & J_h = F(x_N) + \sum_{k=0}^{N-1} \begin{bmatrix} x_k \\ u_k \end{bmatrix}' Q_h \begin{bmatrix} x_k \\ u_k \end{bmatrix}, \\
s.t. \quad & x_{k+1} = Ax_k + Bu_k, \\
& x_k \in \mathbb{X}, \quad u_k \in \mathbb{U}, \quad x_N \in \mathbb{X}_f, \\
& \forall k \in \{0, 1, \dots, N-1\},
\end{aligned} \tag{3.11}$$

where $F(x_N) = x_N' P_c x_N$, A , B and Q_h are given by (3.5) and (3.8), respectively. $\mathbb{U}$ is a convex, polyhedral and compact subset of $\mathbb{R}^m$ and $\mathbb{X}$ a convex, polyhedral and closed subset of $\mathbb{R}^n$, each set containing the origin in its interior.

As shown in the Chapter 2 the recursive feasibility of the MPC (3.11) is guaranteed by ensuring the initial state is inside the initial feasible set which is defined in Definition 2.2.

REMARK 3.1 *The MPC (3.11) includes a cross term of state and input. This difference from the MPC (1.3) results that K_f in the design of terminal invariant set of MPC (1.3), i.e., (2.1) becomes*

$$K_f = -(R_0 + B' P_h B)^{-1} (B' P_h A + S_0') \tag{3.12}$$

where P_h is the stationary solution of the following algebraic Riccati equation

$$P_h = A' P_h A - (A' P_h B + S_0) (B' P_h B + R_0)^{-1} (B' P_h A + S_0') + Q_0. \tag{3.13}$$

3.3 Relationship between Initial feasible set of MPC and S/H time

Because the constraints $\mathbb{X}$ and $\mathbb{U}$ are H-polytopes the initial feasible set will be also an H-polytope [21]. From definition 2.2 the initial feasible set is related with the prediction horizon N and the terminal invariant set $\mathbb{X}_f$. While the prediction horizon N is related to the S/H time according to the equation (3.10) and the terminal invariant set $\mathbb{X}_f$ is also related to the S/H time according to definition 2.1, the initial feasible set is related to the S/H time. However, the relationship between the initial feasible set and the S/H time has not been investigated. In the rest of the section we illustrate the relationship between the initial feasible set and the S/H time through simulations of the second order system with the following transfer function

$$G(s) = \frac{p_1 p_2}{(s - p_1)(s - p_2)}, \quad (3.14)$$

where p_1 and p_2 are the two poles of the system. Here the controllable canonical form of the system (3.14) is used in the design of MPC. The parameters of MPC are $T = 2$ s, $R = 0.1$, $Q = I_2$. The constraints are

$$\begin{bmatrix} -1 \\ -1 \end{bmatrix} \leq x_k \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad -0.5 \leq u_k \leq 0.5. \quad (3.15)$$

3.3.1 Stable system

The second order system (3.14) is stable when $\text{Re}(p_1) \leq 0$ and $\text{Re}(p_2) \leq 0$. In general the poles of the system (3.14) have the following three stable cases:

1. Two distinct real stable poles;
2. Two equal real stable poles;

3. Two stable complex poles.

Figure 3.1: Stable system with two distinct real roots $p_1 = -6$, $p_2 = -4$: (a) size (area) of initial feasible set vs h ; (b) the initial feasible sets. Note that the initial feasible set with $h = 0.4$ s includes the one with $h = 0.25$ s while the initial feasible set with $h = 0.25$ s doesn't include the one with $h = 0.2$ s.

Figure 3.1 shows that the size (area) of the initial feasible set increases as the S/H time increasing for the stable case 1. For example, for $h = 0.5$ s and $h = 1$ s, the same continuous time cost function (3.2) is discretized properly which gives (3.6). Then different K_f are determined by (3.12) which result in different terminal invariant sets as shown in Figure 3.2. As shown in Figure 3.2 the terminal invariant set for $h = 1$ s is bigger than the one for $h = 0.5$ s. The discrete time horizon is determined by

$$N = \frac{T}{h}$$

which is different for $h = 0.5$ s and $h = 1$ s. The initial feasible sets (as shown in Figure 3.3) are then calculated from the terminal invariant sets. Figure 3.3 shows that the initial feasible set for $h = 1$ s is bigger than the one for $h = 0.5$ s.

Figure 3.2: Terminal invariant sets of $h = 0.5$ s and $h = 1$ s.

Figure 3.3: Initial feasible sets of $h = 0.5$ s and $h = 1$ s.

In Figure 3.1, note that the initial feasible set with a bigger size doesn't always include the smaller one (e.g. in Figure 3.1 the initial feasible set with $h = 0.4$ s and area 3.082 includes the one with $h = 0.25$ s and area 2.386 while the initial feasible set with $h = 0.25$ s and area 2.386 doesn't include the one with $h = 0.2$ s and area 2.334). Also we can see that the size of the initial feasible set remains almost the same when h is small. The reason is that there exists a finite S/H time such that the achieved performance is arbitrarily closer to the continuous time performance [109]. Similar observations can be made when the system has equal stable or two stable complex poles.

3.3.2 Unstable system

If the second order system (3.14) is unstable the poles of the system have the following four cases:

1. Two distinct real unstable poles;
2. Two equal real unstable poles;
3. Two unstable complex poles;
4. one stable pole and one unstable pole.

Figure 3.4 shows that for the unstable case 1 the size of the initial feasible set decreases when the S/H time increases. This is expected because for unstable system, performance deteriorates as the S/H time increases. Also we can see that the initial feasible set with bigger size doesn't always include the one with smaller size for unstable system (e.g. in Figure 3.4 the initial feasible set with $h = 0.2$ s and area 0.0075 does not include the one with $h = 0.25$ s and area 0.0071). Similar observations can be made when the system has equal unstable or two unstable complex

poles. Figure 3.5 shows that the size of the initial feasible set are almost the same when the S/H time changes for the unstable case 4.

From this simple example we find that the relationship between the initial feasible set and the S/H time is complicated and there is no general pattern to follow. In next section an example is given to illustrate how to select S/H time, taking into account the initial feasible set, closed-loop performance, as well as computational complexity.

Figure 3.4: Unstable system with two distinct real roots $p_1 = 5$, $p_2 = 3$: (a) size (area) of initial feasible set vs h ; (b) the initial feasible sets. Note that the initial feasible set with $h = 0.2$ s does not include the one with $h = 0.25$ s.

3.4 Illustrative Example

In this section we show, by simulation example, the relationships between closed-loop cost, the size of initial feasible set, the computational time of MPC and S/H time for the constrained MPC problem (3.11).

Figure 3.5: Unstable system with one stable root and one unstable root $p_1 = -6$, $p_2 = 3$: (a) size (area) of initial feasible set vs h ; (b) the initial feasible sets. Note that the size of the initial feasible set are almost the same when the S/H time changes.

Consider the continuous-time system (1.4) where

$$A_c = \begin{bmatrix} 11 & -30 \\ 1 & 0 \end{bmatrix}, \quad B_c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C_c = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

The eigenvalues of the above system are $p_1 = 6$ and $p_2 = 5$. The cost function is given by (3.2) with $Q = I_2$, $R = 0.1$ and $T = 2$ s. To evaluate the performance with S/H time h changing, the system is driven to the origin from the initial condition $x_0 = [0.1 \ 0]'$. The closed-loop cost is calculated using the closed-loop trajectory in the simulation and defined as

$$F = \sum_{k=0}^{Nsim} \begin{bmatrix} x_k \\ u_k \end{bmatrix}' Q_h \begin{bmatrix} x_k \\ u_k \end{bmatrix}, \quad (3.16)$$

where $Nsim = \lceil \frac{T_{sim}}{h} \rceil$ and $T_{sim} = 3$ s is the simulation time. We assume the

following constraints on the system

$$\begin{bmatrix} -1 \\ -1 \end{bmatrix} \leq x_k \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad -1 \leq u_k \leq 2.5.$$

We use MATLAB built-in solver `quadprog` to solve the QP problem. The computation is performed on a desktop PC with a 2.80 GHz CPU, 4 GB of RAM and running MATLAB's version 7.12.0.635.

Figure 3.6(a) shows the closed-loop costs versus S/H time h and confirms that larger S/H time h leads to bigger closed-loop cost (i.e., poorer closed-loop performance). In Figure 3.6(b) the size (area) of initial feasible sets of MPC with different S/H time are shown. Just as shown in the Section 3.2 the area of the initial feasible set for the unstable system is decreasing as the S/H time increasing. Although small S/H time is preferred for larger initial feasible set and lower closed-loop cost, smaller S/H time would require more computational resources shown in Figure 3.7.

An appropriate balance between performance and implementation cost is crucial especially when resources are limited. For example, from Figures 3.6(a) and 3.6(b), reducing S/H time h from 0.4 s to 0.25 s (8% increase in implementation cost, assuming implementation cost is proportional to computational time of MPC) improves performance by 96%, while increasing the size of the initial feasible set by 44% as shown in Figure 3.8. However, a similar investment in computational resources to further reduce S/H time h from 0.25 s to 0.2 s (a 15% increase in implementation cost) only gives a 2% improvement in performance and 14% increase in the size of initial feasible set. So further investment in computational resources may not be worthy. Thus the 0.25 s is an appropriate selection in the trade-off between their on-line computational complexity, initial feasible set, and closed-loop performance.

(a)

(b)

Figure 3.6: (a) Closed-loop cost vs S/H time h ; (b) Area of initial feasible set vs S/H time h .

Figure 3.7: Computational time vs S/H time.

Figure 3.8: Initial feasible set of the MPC when $h = 0.2$ s, $h = 0.25$ s and $h = 0.4$ s.

3.5 Conclusion

In this chapter, the effect of S/H time on the initial feasible set in MPC is investigated. In our study, although we have adopted a continuous time cost function to account for the discretization effect when S/H time changes, the size and shape of the initial feasible set still depends in a complicated manner with the S/H time. Recognizing that many other factors, such as closed-loop performance, computational load could affect the choice of S/H time, an example is given to illustrate the selection of S/H time for the case when trade-off between implementation cost and performance benefits is required.

Chapter 4

A Trajectory Advisor for Model Predictive Control with Tracking Error Bound and a Measurement Checker for Moving Horizon Estimation

In this chapter we showed the relationship between a trajectory advisor for MPC with tracking error bound and a measurement checker for constrained MHE. Akin to the concept of reference governor, the trajectory advisor can supply a new feasible reference signal to recover the feasibility of the MPC problem if the original reference signal makes the problem infeasible. In addition, from symmetry, an equivalent constrained MHE can be formulated. Then the counterpart of the trajectory advisor for MPC is a measurement checker for constrained MHE. The measurement checker could be used to determine the admissibility of the measurements before the constrained MHE is solved. Some subtle differences between MPC and MHE are

highlighted in this chapter.

4.1 Introduction

MPC [75, 92] and MHE [43, 92] are both optimization-based methods where a constrained optimization problem is solved at each time instant. In process control, constraints are often imposed on the control input, such as the valves (completely closed and completely opened), and system state, such as the temperature in a reaction vessel (minimal and maximal reaction temperature). MPC solves an online constrained optimization to obtain an admissible input sequence that ensures the state constraints are satisfied. For estimation, constraints could also be imposed on process and measurement noises to model non-Gaussian noise, such as truncated Gaussian noise [43] and mixture of two Gaussian noises [94], thus constraining the estimator to give only estimates that are within a priori known limits.

The relationship between control and estimation has been concerned a lot in the literature. The Lagrangian duality between LQR and Kalman filter for unconstrained linear system is well understood [59]. For linear system with constraints, the Lagrangian duality between constrained control and constrained estimation has also been investigated [42]. In addition, the symmetry between LQT problem and Kalman filter for unconstrained continuous time linear system and constrained discrete time linear system has been discussed in [100] and [76] respectively. In this chapter, by leveraging the symmetry exposed in [76], we showed relationship of a trajectory advisor of the MPC with tracking error bound and a measurement checker of constrained MHE.

Recursive feasibility of the constrained optimization problem plays a key role in MPC and MHE. For MPC, recursive feasibility ensures that the state is not driven into a "blind alley" where the constrained problem could become infeasible. For

MHE, the constrained optimization problem could become infeasible when erroneous measurements occur. For constrained tracking problem, reference governor was originally designed to guarantee its feasibility [40]. The basic idea of reference governor is to add an auxiliary nonlinear device between the reference command input and the closed-loop control system. The auxiliary nonlinear device on-line adjusts the trajectory to be followed. Recently several MPC schemes have been proposed to guarantee the feasibility of constrained tracking problem [70, 71, 82]. In [71] admissible steady state and input are introduced as additional decision variables in a modified cost function of MPC. The control law ensures feasibility for piecewise constant reference. A robust version of [71] which could handle additive disturbance by using tube-based MPC is investigated in [70]. In [82] the maximum invariant set of the feasible reference signal and the current system state is determined. A reference governor based on the invariant set is proposed to recover the feasibility of the MPC tracking problem. Given the current system state the reference governor could supply a new feasible reference signal to the MPC if the original one makes the MPC infeasible. In this chapter, the initial feasible set of a standard MPC tracking problem with tracking error bound is determined. Based on the feasible set, the trajectory advisor is constructed to recover the feasibility of the MPC by supplying a feasible reference to MPC if the original reference makes the MPC infeasible. We show in simulations the performance of MPC with trajectory advisor. In addition, we highlight that to achieve offset free tracking in the presence of unknown constant disturbance, incremental model should be used and the cost function suitably modified. The potential application of the trajectory advisor is dual-stage automated tooling machines [23, 45] where the difference between the reference signal and the trajectory of the slow stage is required to be within a range.

For constrained MHE, erroneous measurements can make the constrained optimization in MHE infeasible. By symmetry, the counter part of the trajectory advi-

sor for the MPC is the measurement checker for the constrained MHE. We show the relationship between the trajectory advisor for MPC with tracking error bound and the measurement checker for constrained MHE. The measurement checker could be used to determine the feasibility of the MHE before the actual constrained optimization is carried out.

The chapter is organized as follows. Section 4.2 formulates the MPC with tracking error bound and gives the condition for recursive feasibility. In Section 4.3 an algorithm for trajectory advisor for the MPC is presented. Section 4.4 discusses the MHE problem formulation through the lens of MPC-MHE symmetry. Based on the symmetry, a measurement checker is constructed for the MHE. In Section 4.5, examples are given to illustrate the MPC and its accompanied trajectory advisor, as well as the measurement checker for MHE. Section 4.6 concludes this chapter.

4.2 Problem Formulation — MPC with tracking error bound

Consider the following linear system

$$\begin{aligned}x_{k+1} &= Ax_k + Bu_k, \\y_k &= Cx_k, \\e_k &= y_k^r - Cx_k,\end{aligned}\tag{4.1}$$

where $y_k \in \mathbb{R}^p$ is the output, y_k^r is the reference signal in the set Ω_r , (A,B) is stabilizable and e_k is the tracking error. The input u_k is constrained to be in Ω_u . We aim at tracking the reference signal y_k^r with the input constraint Ω_u and tracking error bound Ω_e satisfied all the time. Assume Ω_r , Ω_u and Ω_e are H-polytopes containing the origin.

The standard MPC with predefined tracking error bounds is defined as the following optimization problem:

$$\begin{aligned}
\min_{\{u_k\}_{k=0}^{N-1}} J(x_0) &= \frac{1}{2} \sum_{k=0}^{N-1} (\|u_k\|_R^2 + \|y_k^r - Cx_k\|_Q^2) + \frac{1}{2} \|y_N^r - y_N\|_{P_N}^2, \quad (4.2) \\
\text{s.t. } x_{k+1} &= Ax_k + Bu_k, \quad (k = 0, 1, \dots, N-1) \\
e_k &= y_k^r - Cx_k \in \Omega_e, \quad (k = 0, 1, \dots, N) \\
u_k &\in \Omega_u, \quad (k = 0, 1, \dots, N-1) \\
x_N &\in \Omega_N,
\end{aligned}$$

where Ω_N is a terminal invariant set and x_k denotes the predicted state vector obtained by applying the input sequence $\{u_i\}_{i=0}^{k-1}$ to the system (4.1) starting at x_0 , $P_N \succeq 0$, $Q \succ 0$, $R \succ 0$ and $(A, C\sqrt{Q})$ is observable. Here we assume that current state x_0 is known and the reference signal y_k^r is available in advance for the prediction window. The solution of MPC (4.2) is denoted as $\{u_k^*\}_{k=0}^{N-1}$.

THEOREM 4.1 RECURSIVE FEASIBILITY OF THE MPC (4.2).

Given that the MPC (4.2) is initially feasible, then problem (4.2) is recursively feasible if $\Omega_N = \Pi_n^1(\Omega_N^t)$ where Ω_N^t is a terminal invariant set for the following augmented system

$$x_{k+1}^t = A^t x_k^t + B^t u_k + B^n \eta_k \quad (4.3)$$

where

$$x_k^t = \left[x_k' \mid (y_k^r)' \quad (y_{k+1}^r)' \quad \cdots \quad (y_{k+N}^r)' \right]', \quad \eta_k = y_{k+N+1}^r \in \Omega_r,$$

$$A^t = \left[\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{0} & A^r \end{array} \right] = \left[\begin{array}{c|cccccc} A & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & I_p & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_p & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & I_p \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right], \quad B^t = \begin{bmatrix} B \\ \mathbf{0} \end{bmatrix}, \quad B^\eta = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ I_p \end{bmatrix}$$

when a pre-specified stabilizing state-feedback controller

$$u_k = -Kx_k^t$$

is applied, namely

$$x_k^t \in \Omega_N^t \Rightarrow -Kx_k^t \in \Omega_u \text{ and } (A^t - B^t K) x_k^t + B^\eta \eta_k \in \Omega_N^t, \quad [-C \ C^r] x_k^t \in \Omega_e \text{ for } \forall \eta_k \in \Omega_r, \quad (4.4)$$

$$\text{where } C^r = \begin{bmatrix} I_p & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

Proof. Rewrite (4.2) as the following equivalent MPC regulation problem:

$$\begin{aligned} \min_{\{u_k\}_{k=0}^{N-1}} J^t(x_k^t) &= \frac{1}{2} \sum_{k=0}^{N-1} (\|u_k\|_{R^t}^2 + \|x_k^t\|_{Q^t}^2) + \frac{1}{2} \|x_N^t\|_{P_N^t}^2, \\ \text{s.t. } x_{k+1}^t &= A^t x_k^t + B^t u_k + B^\eta \eta_k, \quad (k = 0, 1, \dots, N-1) \\ C^t x_k^t &\in \Omega_e, \quad \eta_k \in \Omega_r, \quad (k = 0, 1, \dots, N) \\ u_k &\in \Omega_u, \quad (k = 0, 1, \dots, N-1) \\ x_N^t &\in \Omega_N^t, \end{aligned} \quad (4.5)$$

where

$$C^t = \left[-C \mid C^r \right], \quad P_N^t = (C^t)' P_N C^t, \quad Q^t = (C^t)' Q C^t, \quad R^t = R. \quad (4.6)$$

According to Definition 2.2 the initial feasible set of the system (4.3) is

$$\begin{aligned} \Omega_0^t(N, \Omega_N^t) = \{ & x_0^t \in \mathbb{R}^{n+p(N+1)} : \exists \{u_k \in \Omega_u, \eta_k \in \Omega_r\}_0^{N-1}, \text{ such that} \\ & \{e_k \in \Omega_e\}_0^N \text{ and } x_N^t \in \Omega_N^t\}. \end{aligned} \quad (4.7)$$

The stabilisability of (A, B) implies that (A^t, B^t) is stabilizable [17]; $P_N \succeq 0$, $Q \succ 0$ and $R \succ 0$ gives $P_N^t \succeq 0$, $Q^t \succ 0$ and $R^t \succ 0$. Together with the observability of $(A, C\sqrt{Q})$, the MPC regulation problem (4.5) is recursively feasible if the initial state of the augmented system is included in the initial feasible set $\Omega_0^t(N, \Omega_N^t)$ [78].

Therefore, by setting

$$\Omega_N = \Pi_n^1(\Omega_N^t), \quad (4.8)$$

the MPC (4.2) is recursively feasible if the initial state and the initial $N+1$ reference signals are in $\Omega_0^t(N, \Omega_N^t)$. $\square$

REMARK 4.1 *Since the constraints Ω_e , Ω_u and Ω_r are H -polytopes containing the origin and with the affine properties of the augmented state model (4.3), it is clear that both the initial feasible set $\Omega_0^t(N, \Omega_N^t)$ and terminal invariant set Ω_N^t are H -polytopes. Here we denote*

$$\Omega_0^t(N, \Omega_N^t) = \left\{ x_0^t \in \mathbb{R}^{n+p(N+1)} : \begin{bmatrix} H_1 & H_2 & H_3 \end{bmatrix} \begin{bmatrix} x_0 \\ Y_0^r \\ y_N^r \end{bmatrix} \leq f \right\}, \quad (4.9)$$

where $Y_0^r = \begin{bmatrix} (y_0^r)' & (y_1^r)' & \cdots & (y_{N-1}^r)' \end{bmatrix}'$ and H_1, H_2, H_3 are matrices of appropriate dimensions.

4.3 Trajectory Advisor for MPC with tracking error bound

From Theorem 4.1, in order to guarantee the recursive feasibility of the MPC (4.2), the current state x_0 and the reference signals $\{y_0^r, y_1^r, \dots, y_N^r\}$ must lie in the set $\Omega_0^t(N, \Omega_N^t)$. In literature, there are at least two reference updating schemes. One is to update the entire reference sequence over the prediction horizon while the other is to adopt the receding horizon strategy to update y_N^r and discard y_0^r . In this chapter we adopt the receding horizon strategy to update the reference. With x_0 and Y_0^r given, the remaining degree of freedom to ensure the feasibility of problem (4.2) is to adjust y_N^r . Thus, we proposed the trajectory advisor for the MPC in Algorithm 1:

TRAJECTORY ADVISOR

Given the current state x_0 and the reference signals $Y_0^r = \begin{bmatrix} (y_0^r)' & (y_1^r)' & \dots & (y_{N-1}^r)' \end{bmatrix}'$,

1. Compute the feasible set $\Omega_{y_N^r}$ by slicing the set $\Omega_0^t(N, \Omega_N^t)$ through x_0 and Y_0^r , i.e.,

$$\Omega_{y_N^r} = \left\{ y_N^r \in \mathbb{R}^p : H_3 y_N^r \leq f - \begin{bmatrix} H_1 & H_2 \end{bmatrix} \begin{bmatrix} x_0 \\ Y_0^r \end{bmatrix} \right\}. \quad (4.10)$$

2. Solve the following constrained optimization problem

$$\begin{aligned} y_N^{r*} &= \arg \min_{y^r} \|y_N^r - y^r\|^2, \\ \text{s.t.} \quad &y^r \in \Omega_{y_N^r}, \end{aligned} \quad (4.11)$$

3. Supply the new feasible reference signal y_N^{r*} to the MPC (4.2).
-

According to Theorem 4.1 and the definition of augmented initial feasible set $\Omega_0^t(N, \Omega_N^t)$, i.e., (4.7), the MPC problem is always feasible if the augmented state is in $\Omega_0^t(N, \Omega_N^t)$. In the algorithm of trajectory advisor, i.e., (4.10)-(4.11), the new reference is selected such that the system state and the reference signals are in

$\Omega_0^t(N, \Omega_N^t)$. So the trajectory advisor can guarantee the feasibility of MPC problem.

REMARK 4.2 *According to Theorem 2.1, the initial feasible set $\Omega_0^t(N, \Omega_N^t)$ is a positively invariant set. Thus $\Omega_0^t(N, \Omega_N^t)$ can be determined offline and computed from the terminal set Ω_N^t using algorithm from [60].*

4.4 Measurement Checker for MHE

Before discussing the measurement checker we reviewed the symmetry between MPC and constrained MHE exposed in [76] first.

4.4.1 Symmetry between MPC and MHE

Consider the following observable system

$$\begin{aligned} x_{k+1} &= A_e x_k + B_e \omega_k, \\ y_k &= C_e x_k, \\ v_k &= y_k^d - C_e x_k, \end{aligned} \tag{4.12}$$

where $y_k^d \in \mathbb{R}^p$ is the measurement, $y_k \in \mathbb{R}^p$ is the output and $x_k \in \mathbb{R}^n$ is the state. The pair (C_e, A_e) is observable. The process noise ω_k and measurement noise v_k are assumed zero mean and white with variances Q_e and R_e respectively. Assume that the initial state estimate μ_{e0} with the associated variance $P_{e0} \succ 0$ is uncorrelated with the two noise sequences. Given the N measurements $\{y_k^d\}_{k=1}^N$, the constrained MHE is defined as

$$P_N^e : \min_{\hat{x}_N} V_N^*(\hat{x}_N) \tag{4.13a}$$

$$V_N^*(\hat{x}_N) = \min_{\{\hat{w}_k\}_{k=0}^{N-1}} \frac{1}{2} \|\hat{x}_0 - \mu_{e0}\|_{P_{e0}^{-1}}^2 + \frac{1}{2} \sum_{k=0}^{N-1} \|\hat{w}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=1}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2, \quad (4.13b)$$

$$s.t. \quad \hat{x}_{k+1} = A_e \hat{x}_k + B_e \hat{w}_k,$$

$$\hat{v}_k = y_k^d - C_e \hat{x}_k,$$

$$\hat{w}_k \in \Omega_w, \quad \hat{v}_k \in \Omega_v, \quad \hat{x}_0 \in \Omega_{x_0},$$

where Ω_w , Ω_v and Ω_{x_0} are H-polytopes containing the origin. Here the optimization is decomposed to an inner optimization (4.13b) where $\hat{x}_N$ is assumed to be known and $\{\hat{w}_k\}_{k=0}^{N-1}$ are decision variables, and an outer optimization (4.13a) where $\hat{x}_N$ is the decision variable. The purpose of this decomposition is to highlight the symmetry between the MHE problem with the MPC (4.2).

The inner optimization of the MHE problem can be converted to a MPC with tracking error bound through the parameter and constraint sets translations shown in Tables 4.1 and 4.2 (see ref. [76]).

$$\begin{aligned} P_N^c : \quad & \min_{\{u_k\}_{k=0}^{N-1}} \frac{1}{2} \sum_{k=0}^{N-1} (\|u_k\|_{R_c}^2 + \|y_k^r - C_c x_k\|_{Q_c}^2) + \frac{1}{2} \|x_N - \mu_{cN}\|_{P_{cN}}^2, \\ & s.t. \quad x_{k+1} = A_c x_k + B_c u_k, \\ & \quad e_k = y_k^r - C_c x_k, \\ & \quad u_k \in \Omega_u, \quad e_k \in \Omega_e, \quad x_N \in \Omega_{x_N}. \end{aligned} \quad (4.14)$$

Table 4.1: Variables translation (see ref. [76], Table 1)

P_N^c	P_N^e	P_N^c	P_N^e	P_N^c	P_N^e
u_0	$\Leftrightarrow \hat{\omega}_{N-1}$	x_0	$\Leftrightarrow \hat{x}_N$	y_0^r	$\Leftrightarrow y_N^d$
u_1	$\Leftrightarrow \hat{\omega}_{N-2}$	x_1	$\Leftrightarrow \hat{x}_{N-1}$	y_1^r	$\Leftrightarrow y_{N-1}^d$
$\vdots$	$\Leftrightarrow \vdots$	$\vdots$	$\Leftrightarrow \vdots$	$\vdots$	$\Leftrightarrow \vdots$
u_{N-1}	$\Leftrightarrow \hat{\omega}_0$	x_N	$\Leftrightarrow \hat{x}_0$	y_{N-1}^r	$\Leftrightarrow y_1^d$

Table 4.2: Parameters translation (see ref. [76], Table 2)

P_N^c	P_N^e	P_N^c	P_N^e
A_c	$\Leftrightarrow A_e^{-1}$	Q_c	$\Leftrightarrow R_e^{-1}$
B_c	$\Leftrightarrow -A_e^{-1}B_e$	R_c	$\Leftrightarrow Q_e^{-1}$
C_c	$\Leftrightarrow C_e$	Ω_u	$\Leftrightarrow \Omega_\omega$
P_{cN}	$\Leftrightarrow P_{e0}^{-1}$	Ω_e	$\Leftrightarrow \Omega_v$
μ_{cN}	$\Leftrightarrow \mu_{e0}$	Ω_{x_N}	$\Leftrightarrow \Omega_{x_0}$

Figure 4.1 further illustrates the symmetry between MPC and MHE where we highlight the symmetry between the set $\Omega_{\hat{x}_N, y_k^d}(N)$, and $\Pi_{n+Np}^1(\Omega_0^t(N, \Omega_N^t))$ where

$$\Omega_{\hat{x}_N, y_k^d}(N) = \left\{ \begin{bmatrix} \hat{x}_N \\ y_N^d \\ \vdots \\ y_1^d \end{bmatrix} \in \mathbb{R}^{n+pN} : G \begin{bmatrix} \hat{x}_N \\ y_N^d \\ \vdots \\ y_1^d \end{bmatrix} \leq b \right\}. \quad (4.15)$$

For MPC with tracking error bound, the set $\Pi_{n+Np}^1(\Omega_0^t(N, \Omega_N^t))$ indicates the feasibility condition of the MPC. Thus, by the symmetry, the set $\Omega_{\hat{x}_N, y_k^d}(N)$ indicates

Figure 4.1: Symmetry between the MHE (4.13) and the MPC with tracking error bound (4.14) .

the feasibility condition of the MHE.

REMARK 4.3 Unlike the MPC in (4.2), there is no requirement that the set Ω_{x_N} in (4.14) be invariant. The Ω_{x_N} in (4.14) is obtained through parameter translation shown in Table 4.2. In MPC, terminal invariant constraint is needed to ensure recursive feasibility. In MHE, the initial constraint Ω_{x_0} represents the prior knowledge of the initial state.

4.4.2 Measurement Checker for MHE

The set $\{y_N^d, y_{N-1}^d, y_1^d\}$ in (4.15) is the set of admissible measurements for MHE. It can be computed by projecting $\Omega_{\hat{x}_N, y_k^d}(N)$ onto the last $N \times p$ coordinates, i.e., $\Omega_{y_k^d} = \Pi_{Np}^2 \left(\Omega_{\hat{x}_N, y_k^d}(N) \right)$. Thus, the following measurement checker is proposed:

MEASUREMENT CHECKER FOR CONSTRAINED MHE (4.13)

Whenever new measurement y_N^d is available,

1. Check whether the measurements $\{y_N^d, \dots, y_1^d\}$ is inside $\Omega_{y_k^d}$.
 2. If the answer is "yes", solve the MHE problem; if the answer is "no", output "the set of measurements $\{y_N^d, \dots, y_1^d\}$ is not admissible".
-

REMARK 4.4 *The set $\Omega_{y_k^d}$ can be computed offline. The online computation of measurement checker is to check whether the N measurements in the moving window are in $\Omega_{y_k^d}$ or not. The computational complexity of this checking is $\mathcal{O}(n)$.*

The measurement checker is only used to find the “not admissible” measurements and does not affect the feasibility of MHE. If the erroneous measurement is found, one method is to discard the erroneous measurement. More specifically, the erroneous measurement is excluded from the cost function of MHE (4.13). Another method is to force the erroneous measurement into the admissible set $\Omega_{y_k^d}$.

REMARK 4.5 *In MPC, the current state x_0 is given while in MHE $\hat{x}_N$ is a decision variable in the constrained optimization problem. Thus unlike the trajectory advisor where the feasible set of reference signal is obtained by slicing the initial feasible set of the MPC, the set of admissible measurements for MHE is obtained by projecting $\Omega_{\hat{x}_N, y_k^d}(N)$ onto $\{y_N^d, y_{N-1}^d, \dots, y_1^d\}$.*

4.5 Illustrative Examples

In this section we give three examples to illustrate the performance of the MPC, the behavior of the trajectory advisor and the measurement checker. In the first example we show the behavior of the trajectory advisor, the performance improvement of the MPC by including tracking error bound and that the offset can be eliminated by using the incremental model of the system in the design of MPC. The second example shows the application of the MPC with trajectory advisor in a dual-stage automated tooling machine. The third example illustrates the measurement checker for MHE.

4.5.1 MPC with Trajectory Advisor

Consider the following system with a constant disturbance $d_k = 0.5$ on the system output

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 1 & 0.3 \\ -0.2 & 0.8 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} u_k \\ y_k &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + d_k \end{aligned} \quad (4.16)$$

The weighting matrices and prediction horizon are

$$Q = 0.001, \quad R = 0.25, \quad P_N = 0.1, \quad N = 3.$$

The constraints on the input u_k , the input increments $\Delta u_k = u_k - u_{k-1}$, the tracking error e_k and reference signal y_k^r are:

$$-0.5 \leq u_k \leq 1, \quad -0.3 \leq \Delta u_k \leq 0.3, \quad -3 \leq e_k \leq 3, \quad -1 \leq y_k^r \leq 2.$$

and the initial state is $x_0 = \begin{bmatrix} -3 & -2 \end{bmatrix}'$.

Nonzero offset

The system model (4.16) is directly used in the design of the MPC (4.2). The system output y_k , the original reference y_k^r , trajectory advisor output y_k^{r*} , the feasible reference set $\Omega_{y_N^r}$ (the vertical red line) and tracking error of the MPC with and without tracking error bound are shown in Figure 4.2. The corresponding system inputs and input incremental are shown in Figure 4.3. In Figure 4.2 the MPC gives nonzero offset in the presence of an unknown constant disturbance. However, at the beginning, with the help of the trajectory advisor, the MPC with tracking error bound converges faster to a steady state value which is closer to the reference and

guarantees the tracking error bound. At $k = 51, 52, \dots, 70$, the original reference signal is infeasible, i.e., $y_k^r \notin \Omega_{y_N^r}$, for both MPC with and without tracking error bound and the trajectory advisor adjusts the reference to a feasible one which is the closest to the original one. Figure 4.3 shows that the constraints of the system inputs and the input incremental are satisfied.

Figure 4.2: Comparison of MPC with and without tracking error bound, output of trajectory advisor y_k^{r*} and the feasible reference set $\Omega_{y_N^r}$ (the vertical red line).

Figure 4.3: System inputs and incremental inputs of the MPC with and without tracking error bound

Zero offset

In order to eliminate the offset the following incremental model is used in the design of the MPC

$$\begin{aligned} \begin{bmatrix} \Delta x_{k+1} \\ y_{k+1} \end{bmatrix} &= \begin{bmatrix} A & \mathbf{0}_{n \times p} \\ CA & I_p \end{bmatrix} \begin{bmatrix} \Delta x_k \\ y_k \end{bmatrix} + \begin{bmatrix} B \\ CB \end{bmatrix} \Delta u_k, \\ y_k &= \begin{bmatrix} \mathbf{0}_{p \times n} & I_p \end{bmatrix} \begin{bmatrix} \Delta x_k \\ y_k \end{bmatrix}. \end{aligned} \quad (4.17)$$

and with u_k replaced by Δu_k in cost function (4.2).

The original reference y_k^r , trajectory advisor output y_k^{r*} , system output y_k , the reference feasible set $\Omega_{y_N^r}$ and tracking error of the MPC with and without tracking error bound are shown in Figure 4.4. The corresponding system input and input incremental are shown in Figure 4.5. As expected, Figure 4.4 shows that zero offset is achieved when incremental model and the modified cost function are used. Again, at the beginning, with the help of the trajectory advisor, the MPC with tracking error bound converges to the steady state faster and guarantees the tracking error bound. At $k = 51, 52, \dots, 70$, the original reference signal is infeasible, i.e., $y_k^r \notin \Omega_{y_N^r}$, for both MPC with and without tracking error and the trajectory advisor adjusts the reference to a feasible one which is the closest to the original one. Figure 4.5 shows that the system inputs and the input incremental within the constraints. In this simulation we adopted integrator augmentation to eliminate steady state error. The integrator augmentation is only a simple method. The internal model principle is a more general approach.

Figure 4.4: With incremental model, comparison of MPC with and without tracking error bound, output of trajectory advisor y_k^{r*} and the feasible reference set $\Omega_{y_N^r}$ (the vertical red line).

Figure 4.5: System input and incremental input of the MPC with and without tracking error bound.

4.5.2 MPC with trajectory advisor for dual-stage automated tooling machine

Consider a dual-stage automated tooling machine with the following continuous model for the x-axis of the slow stage [23, 45].

$$\dot{x}_t = \begin{bmatrix} 0 & 1 \\ 0 & a \end{bmatrix} x_t + \begin{bmatrix} 0 \\ b \end{bmatrix} u_t \quad (4.18)$$

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t \quad (4.19)$$

where the first element of x_t is the position, the second element of x_t is the velocity, $u(t)$ is the input torque and $a = -1.08$, $b = 0.34$. Assume the range of the fast stage is $[-2, 2]$, i.e, the tracking error bound of the slow stage $e \in [-2, 2]$, the reference signal $y_t^r \in [-0.6, 0.6]$ and the input $u_t \in [-14, 14]$. The discrete time system model is obtained by sampling (4.18) with period $h = 1$ s. The performance of the MPC with trajectory advisor are shown in Figures 4.6, 4.7 and 4.8. As shown in Figure

Figure 4.6: The system output y_k , reference signal y_k^r and the output of trajectory advisor y_k^{r*} .

4.6, at the initial time, the trajectory advisor adjusted the reference signal y_k^r (dot-dash cyan line) to feasible reference signal y_k^{r*} (solid blue line) to make sure that

Figure 4.7: Tracking error.

Figure 4.8: The system input.

the tracking error bound is satisfied as shown in Figure 4.7. At $k = 16$ where some unknown disturbance happens, the trajectory advisor adjusted the reference signal y_k^r (dot-dash cyan line) to the feasible reference signal y_k^{r*} (solid blue line) to make the tracking error bound satisfied as shown in Figure 4.7. Figure 4.8 shows that the input constraint is satisfied.

4.5.3 The measurement checker

Consider the system (4.12) with the following matrices:

$$\begin{aligned} A_e &= \begin{bmatrix} 0.8 & 0.4 \\ -0.4 & 0.8 \end{bmatrix}, \quad B_e = \begin{bmatrix} 0.8 \\ 0.3 \end{bmatrix}, \\ C_e &= \begin{bmatrix} 0.5 & 0.4 \end{bmatrix}. \end{aligned}$$

The constraints on process noise, measurement noise and the state are:

$$-3 \leq \omega_k \leq 5, \quad -4 \leq v_k \leq 2, \quad [-3, -3]' \leq x_0 \leq [3, 3]'$$

The parameters in MHE P_N^e are:

$$Q_e = 4, \quad R_e = 2, \quad P_{e0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad N = 3, \quad \mu_{e0} = \begin{bmatrix} 1 & 1 \end{bmatrix}'.$$

Figure 4.9 shows the set of admissible measurements, i.e., $\Omega_{y_k^d}$. The measurement checker proposed in Algorithm 4.4.2 can check the feasibility of the MHE before solving the constrained optimization problem. For example, if the measurements are $\{-4, 2, -8\}$, which is outside the set of admissible measurement as shown in Figure 4.9, the measurement checker will report that this set of measurements is

Figure 4.9: The set of admissible measurements.

not admissible. If the measurements are $\{-4, 2, 3\}$, which is admissible as shown in Figure 4.9, the measurement checker will report that the set of measurement is admissible and the constrained MHE can be solved giving the state estimate $\hat{x}_3^* = \begin{bmatrix} 1.1565 & 1.0543 \end{bmatrix}'$ and noise estimate $\hat{\omega}_0^* = -0.7131$, $\hat{\omega}_1^* = 1.3435$, $\hat{\omega}_2^* = 1.046$.

4.6 Conclusion

In this chapter, we showed the relationship of the trajectory advisor for the standard MPC tracking problem with tracking error bound and the measurement checker for constrained MHE for constrained linear systems. The trajectory advisor is constructed from the initial feasible set of the MPC problem. The trajectory advisor can supply a new feasible reference signal to recover the feasibility of the MPC problem if the original reference signal makes the MPC problem infeasible. The illustrated examples show that by including tracking error bound, MPC gives faster transient than without tracking error bound. Although closed-loop performance depends very much on tuning, the illustrative examples show that by including tracking error bound, the MPC has the potential to give faster transient than without tracking

error bound. Incremental input model and suitable modification of the cost function are needed to achieve zero steady state error in the presence of constant disturbance. The simulation also shows the application of the MPC with trajectory advisor to the dual-stage automated tool machines. From symmetry the counter part of trajectory advisor for the MPC is the measurement checker for the constrained MHE. The measurement checker could be used to determine the admissibility of the measurements before the constrained MHE is solved. The computational complexity of the initial feasible set is not discussed in this thesis. This is our future work.

Chapter 5

Moving Horizon Estimation for ARMAX process with t -Distribution Noise

In this chapter a Maximum Likelihood Estimation (MLE) is proposed for autoregressive-moving-average with exogenous (ARMAX) process with generalized t -distribution (GT) noise assumption. By using the latest N measurements the MLE results a moving or receding horizon estimation which maximizes the likelihood function determined by the noise probability density function (pdf), i.e., the GT distribution pdf. Instead of solving the maximization problem numerically, the Influence Function (IF), an analysis tool in robust statistics, is used to formulate a recursive solution to the maximization problem approximately. It turns out that the resulting recursive estimator has a structure which can be separated into two parts: a non-linear transformation of the measurements and a linear recursive estimator. The thick tail property of GT noise makes the proposed estimator robust to outlier. With the moving horizon or receding horizon structure the proposed estimator is robust to modeling error. As expected, the recursive estimator is reduced to the

well known Moving Window Least-Squares Estimation when the noise is Gaussian noise. A formula for the variance of the estimator is derived and the impact on the estimate is also analyzed.

5.1 Introduction

State estimation is usually required in process control and monitoring, and Gaussian noise is often assumed in state estimation. However, the Gaussian noise assumption is an approximation to reality. The occurrence of outliers, transient data in steady-state measurements, instrument failure, model nonlinearity, etc. can all induce non-Gaussian data [104].

It is well known that t -distribution has the property of thick tail to better model the occurrence of outliers [62]. In addition t -distribution reduces to Gaussian distribution when its shape parameter tends to infinity. Thus t -distribution has the flexibility to characterize noise with Gaussian or non-Gaussian statistical properties. In the literature the heavy tail property of t -distribution has been used to improve the performance of estimators in [1, 2, 79, 95]. These papers are based on Bayesian method and the heavy tail property of t -distribution was used to down weigh the outlier. Our method differs from these estimators in that the probability density function (pdf) of t -distribution was used directly in the likelihood function of Maximum Likelihood Estimation (MLE).

The sensitivity of an estimator when the underlying noise assumption (such as, Gaussian noise) is violated has been extensively studied in robust statistics literature [47, 55]. In particular, Huber [55] studied the effect of outliers by contaminating the underlying noise distribution with data from an arbitrary unknown distribution. Hampel [47] proposed the Influence Function (IF) approach to describe the effect of an infinitesimal contamination of the underlying noise distribution. If the IF of an

estimator is bound and/or decreasing, or increasing slowly for large magnitude of noise, then the estimator is robust to outliers. It can be shown that the IF of least-squares estimator increases linearly with the magnitude of noise and is unbounded [47]. This confirms the well known fact that the standard least-squares estimation is not robust to outliers. Other approaches for handling non-Gaussian noise include particle filter [8] which is based on Monte Carlo method, but the heavy computation load of particle filter limits its applications.

Unlike Hampel, who uses the IF as an analysis tool, the main contribution of this chapter is to employ the IF to formulate a computationally efficient recursive algorithm that gives an approximate solution to the moving horizon MLE for Auto-Regressive-Moving-Average with eXogenous inputs (ARMAX) process with t -distribution noise. Our algorithm reduces to the Kalman filter when the noise is Gaussian and the horizon grows with time instant.

ARMAX process is popularly used in the statistical analysis of time series [10, 20]. The Kalman filter of ARMAX process is well known, see Example 11.6 in [10]. In [53] a MLE of ARMAX process with generalized t -distribution noise was developed. However, in [53] the IF was used to approximate the MLE by using all the measurements. This makes the estimator sluggish when plant parameters changes. In this chapter we derive the moving horizon version of the ARMAX filter. The moving horizon structure makes the proposed estimator to be robust to model mismatch and the moving horizon length N can be used a tuning parameter for the proposed estimator. In addition we derive a computationally efficient recursive algorithm and highlight the connection of recursive algorithms of the moving horizon MLE and ARMAX filter.

The chapter is organized as follows. In Section 5.2 we begin with state space model of the ARMAX process, and use the IF to derive an approximate solution of the MLE for ARMAX process with t -distribution noise. In Section 5.3 we adopt moving

horizon principle, such as in [22, 56, 66, 67, 68, 72], to obtain the moving horizon version of the IF approximation of the MLE and derive a recursive algorithm. The statistical properties of our proposed estimator are analyzed in Section 5.4. The performance of our proposed estimator in the sense of variance of estimate, robustness to outlier and model mismatch is investigated in Section 5.5 by simulations. The proposed MHE is applied two experiments in Section 5.6. Section 5.7 concludes this chapter.

5.2 Maximum Likelihood Estimation for ARMAX process with t -distribution noise

Consider the single-input single-output ARMAX process (2.7) where e_k is assumed as zero mean t -distribution noise with pdf $f(e_k)$ which is represented by [62]

$$f(e_k) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\sigma\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{e_k^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}}, \quad (5.1)$$

where ν is the shape parameter of the distribution, σ is the variance parameter and Γ is the gamma function.

For convenience, we adopt the state space model of the ARMAX process (2.7), as in [10]

$$x_{k+1} = \Phi_a x_k + \Gamma u_k + \Omega e_k, \quad (5.2a)$$

$$y_k = H x_k + e_k, \quad (5.2b)$$

where

$$\Phi_a = \begin{bmatrix} -a_1 & 1 & 0 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ -a_{n-1} & 0 & \cdots & \cdots & 0 & 1 \\ -a_n & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n_B} \\ 0 \end{bmatrix},$$

$$\Omega = \begin{bmatrix} c_1 - a_1 \\ c_2 - a_2 \\ \vdots \\ c_n - a_n \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

where a_i , b_i and c_i are the parameters of ARMAX process defined in (2.7). Given measurements $y_1, y_2, \dots, y_T$, according to the model (5.2) we have

$$y_k = H\Phi^{k-1}x_1 + H\bar{x}_k + e_k \quad \text{for } k = 1, \dots, T, \quad (5.3)$$

where

$$\begin{aligned} \Phi &= \Phi_a - \Omega H, \\ \bar{x}_k &= \sum_{i=1}^{k-1} \Phi^{k-i-1} (\Gamma u_i + \Omega y_i). \end{aligned} \quad (5.4)$$

We denote the MLE for the ARMAX process (5.2) with t -distribution noise as

$$\hat{x}_1 = \arg \min_{x_1} J$$

where

$$J = - \sum_{k=1}^T \ln f(e_k), \quad \text{where } e_k = y_k - H\Phi^{k-1}x_1 - H\bar{x}_k. \quad (5.5)$$

$\hat{x}_1$ can be found by solving the equation

$$\sum_{k=1}^T \psi_k(e_k) = 0, \quad (5.6)$$

where

$$\psi_k(e_k) \triangleq \frac{d}{dx_1} (\ln f(e_k)) = -\frac{(\nu+1) (H\Phi^{k-1})' e_k}{\sigma^2\nu + (e_k)^2}.$$

5.2.1 Influence Function of the MLE for ARMAX process

According to robust statistics literature [47, 108] the solution of (5.6) can be denoted by $T(G_k)$, where T is the functional given by

$$\sum_{k=1}^T \int_{-\infty}^{\infty} \psi_k(y_k, T(G_k)) dG_k(y_k) = 0, \quad (5.7)$$

where G_k and G_k are the cumulative density function (cdf) and the empirical cdf of y_k , respectively. Because in practice, the solution of (5.6) uses empirical measurements we use $T(G_k)$ to denote the solution of (5.6). The IF describes the effect of an infinitesimal contamination at the point y on the estimate, standardized by the mass of the contamination.

In this chapter the noise is assumed to be t distributed with pdf $f(e_k)$, taking the expectation of (5.6) gives

$$\sum_{k=1}^T \int_{-\infty}^{\infty} \psi_k(e_k) f(e_k) de_k = 0. \quad (5.8)$$

Hence to obtain the IF of our estimator, we replace $f(e_k)$ in (5.8) with $(1-h)f(e_k) + h\Delta_{e_k}$

$$\sum_{k=1}^T \int_{-\infty}^{\infty} \psi_k(e_k) [(1-h)f(e_k) + h\Delta_{e_k}] de_k = 0. \quad (5.9)$$

Equation (5.9) implicitly defines our estimator as a function of h and we denote it as $\kappa(h)$. According to Definition 2.8 the IF of our estimator can be obtained by differentiating equation (5.9) wrt. h and setting $h = 0$

$$\text{IF}(e) \triangleq \left. \frac{d\kappa(h)}{dh} \right|_{h=0} = - \left(\sum_{k=1}^T \int_{-\infty}^{\infty} \frac{d\psi_k(e)}{dx_1} f(e) de \right)^{-1} \sum_{k=1}^T \psi_k(e_k) \Big|_{x_1=\bar{x}_1}. \quad (5.10)$$

The derivation of (5.10) is given in Appendix 4. The initial state x_1 is unknown but we assume that its mean $\bar{x}_1$ is known.

5.2.2 IF Approximation of MLE for ARMAX process with t -distribution noise

Since Equation (5.9) implicitly defines our estimator as a function of h and we denote it as $\kappa(h)$. In general, $\kappa(h)$ is a nonlinear state estimator. The Taylor series expansion of $\kappa(h)$ at $h = 0$ is given by

$$\kappa(h) = \bar{x}_1 + \left. \frac{d\kappa(h)}{dh} \right|_{h=0} h + \frac{1}{2} \left. \frac{d^2\kappa(h)}{dh^2} \right|_{h=0} h^2 + \dots \quad (5.11)$$

Since (5.9) reduces to (5.6) when $h = 1$, $\kappa(1)$ is the solution that we are looking for, namely the solution to (5.6). We can get an approximate solution to this by retaining only the first two terms of (5.11). Using the definition of $\text{IF}(e)$ from (5.10), this gives

$$\hat{x}_1 \approx x_{\text{IF}} \triangleq \bar{x}_1 + \text{IF}(e). \quad (5.12)$$

PROPOSITION 5.1 *Given the measurement model represented by (5.3) with the noise e_k having the pdf of (5.1), the IF of MLE for ARMAX process with t -distribution*

noise is

$$\begin{aligned} \text{IF}(e) &= \left(\sum_{k=1}^T \int_{-\infty}^{\infty} (H\Phi^{k-1})' H\Phi^{k-1} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \\ &\quad \times \sum_{k=1}^T \frac{(\nu+1)(H\Phi^{k-1})' e_k}{\sigma^2\nu + (e_k)^2} \Big|_{x_1=\bar{x}_1}. \end{aligned} \quad (5.13)$$

Proof. See Appendix 5. □

COROLLARY 5.1 *When the noise e_k is Gaussian, i.e., e_k having the pdf (5.1) with $\nu = \infty$, (5.13) becomes*

$$\text{IF}(e) = -\bar{x}_1 + \left(\sum_{k=1}^T (H\Phi^{k-1})' H\Phi^{k-1} \right)^{-1} \sum_{k=1}^T (H\Phi^{k-1})' (y_k - H\bar{x}_k) \quad (5.14)$$

and (5.12) becomes

$$\hat{x}_1 = \bar{x}_1 + \text{IF}(e) = \left(\sum_{k=1}^T (H\Phi^{k-1})' H\Phi^{k-1} \right)^{-1} \sum_{k=1}^T (H\Phi^{k-1})' (y_k - H\bar{x}_k) \quad (5.15)$$

which is the least-squares estimation.

Proof. See Appendix 6. □

REMARK 5.1 *It is re-assuring that our estimator reduces to the well known result when $\nu \rightarrow \infty$, and the IF approximation is exactly the least-squares estimation. This is because the pdf $f(e_k)$ is reduced to Gaussian pdf when $\nu \rightarrow \infty$.*

In the sequel, for simplicity, we shall assume that $\bar{x}_1 = 0$. Thus (5.13) is now

$$\begin{aligned} \text{IF}(e) &= \left(\sum_{k=1}^T \int_{-\infty}^{\infty} (H\Phi^{k-1})' H\Phi^{k-1} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \\ &\quad \times \sum_{k=1}^T \frac{(\nu+1)(H\Phi^{k-1})' (y_k - H\bar{x}_k)}{\sigma^2\nu + (y_k - H\bar{x}_k)^2} \end{aligned}$$

$$(5.16)$$

and (5.12) becomes

$$x_{\text{IF}} = \text{IF}(e). \quad (5.17)$$

5.3 Moving Horizon Estimation of ARMAX process with t -distribution noise (MHE-TD)

In this section we adopt the moving horizon principle to obtain the moving horizon version of the estimator. Instead of using all the measurements up to T , using the most recent N measurements $\{y_k\}_{T-N+1}^T$, therefore, for (5.16) replacing $k = 1$ with $k = T - N + 1$ gives the IF of MHE-TD:

$$\begin{aligned} \text{IF}(e) &= \left(\sum_{k=T-N+1}^T (H\Phi^{k-1})' H\Phi^{k-1} \int_{-\infty}^{\infty} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \\ &\times \sum_{k=T-N+1}^T \frac{(\nu+1)(H\Phi^{k-1})'(y_k - H\bar{x}_k)}{\sigma^2\nu + (y_k - H\bar{x}_k)^2}. \end{aligned} \quad (5.18)$$

PROPOSITION 5.2 IF APPROXIMATION OF MHE-TD Given the ARMAX process (2.7) where $C(q^{-1})$ is stable and noise e_k with t -distribution pdf of (5.1) IF approximation of MHE-TD is

$$\hat{x}_T = \bar{x}_T + \left(\sum_{i=1}^N (H\Phi^{i-N})' H\Phi^{i-N} \right)^{-1} \Psi \mathcal{W}_{T,N}, \quad (5.19a)$$

$$\hat{y}_T = H\hat{x}_T, \quad (5.19b)$$

where

$$\bar{x}_T = \sum_{k=1}^{T-1} \Phi^{T-k-1} (\Gamma u_k + \Omega y_k), \quad (5.20a)$$

$$\Psi = \begin{bmatrix} (H\Phi^{1-N})' & (H\Phi^{2-N})' & \cdots & H' \end{bmatrix}, \quad (5.20b)$$

$$\mathcal{W}_{T,N} = \begin{bmatrix} w_{T-N+1} & w_{T-N+2} & \cdots & w_T \end{bmatrix}', \quad (5.20c)$$

$$w_{T-N+i} = \left(\int_{-\infty}^{\infty} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \frac{(\nu+1)(y_{T-N+i} - H\bar{x}_{T-N+i})}{\sigma^2\nu + (y_{T-N+i} - H\bar{x}_{T-N+i})^2},$$

$$i = 1, 2, \dots, N. \quad (5.20d)$$

Proof. See Appendix 7. □

REMARK 5.2 According to Corollary 5.1 when the pdf of noise e_k (5.1) is Gaussian, the IF approximation of MHE-TD, i.e., (5.19a) becomes the following well known Moving Window Least-Squares Estimation with window length N , for the ARMAX process (2.7)

$$\hat{x}_T = \left(\sum_{i=1}^N (H\Phi^{i-N})' H\Phi^{i-N} \right)^{-1} \times \sum_{i=1}^N (H\Phi^{i-N})' (y_{T-N+i} - H\Phi^{i-N} (\Xi_u U_{T-1,N-1} + \Xi_y Y_{T-1,N-1})) \quad (5.21)$$

where

$$U_{T-1,N-1} = \begin{bmatrix} u_{T-N+1} & u_{T-N+2} & \cdots & u_{T-1} \end{bmatrix}',$$

$$Y_{T-1,N-1} = \begin{bmatrix} y_{T-N+1} & y_{T-N+2} & \cdots & y_{T-1} \end{bmatrix}',$$

$$\Xi_u = \begin{bmatrix} \Phi^{N-2}\Gamma & \Phi^{N-3}\Gamma & \cdots & \Gamma \end{bmatrix}, \quad \Xi_y = \begin{bmatrix} \Phi^{N-2}\Omega & \Phi^{N-3}\Omega & \cdots & \Omega \end{bmatrix}.$$

When $\Omega = 0$, $\Gamma = 0$, Equation (5.21) reduces to the least-squares solution (13.5) in [10].

Consider the following ARMAX model

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k \quad (5.22)$$

where $A(q^{-1}) = 1 - 0.8q^{-1}$, $B(q^{-1}) = 0.2q^{-1}$ and $C(q^{-1}) = 1 - 0.6q^{-1}$ which is stable. The state space model of (5.22) is given by

$$x_{k+1} = 0.6x_k + 0.2u_k + 0.2y_k \quad (5.23a)$$

$$y_k = x_k + e_k \quad (5.23b)$$

Suppose the mean of the initial state is 0, i.e., $\bar{x}_0 = 0$. The noise e_k is t -distribution noise with $p = 2$, $q = 1.5$, $\sigma = 0.2\sqrt{2}$ and $u_k = 2$.

Figure 5.1: x_T and $\hat{x}_T$ of IF approximation of MHE-TD.

Figure 5.1 shows the state x_k and estimation of IF approximation of MHE-TD with correct mean of initial state and wrong mean of initial state ($\bar{x}_0 = -3$). From Figure 5.1 we can see that the estimation of IF approximation of MHE-TD with wrong mean of initial state converges to the estimation with correct mean of initial state because the effect of initial condition will disappear after the transient for the stable system (5.23). In this chapter we assumed that $C(q^{-1})$ is stable. So the

wrong initial condition will disappear. Thus for simplicity we assume $\bar{x}_1 = 0$.

In many cases the observations are obtained sequentially. Hence it is desirable to use the latest measurement to compute the current estimation as soon as the new observation is available. The recursive formulation of IF approximation of MHE-TD is given by the following theorem:

THEOREM 5.1 *THE RECURSIVE MHE-TD IF approximation of MHE-TD satisfies the following recursive algorithm*

$$\begin{aligned}\hat{x}_{T+1} &= \Phi\hat{x}_T + \Gamma u_T + \Omega y_T + L_N (z_{T+1} - H(\Phi\hat{x}_T + \Gamma u_T + \Omega y_T)) \\ &\quad - \tilde{L}_N (z_{T-N+1} - H\Phi^{-N+1}\hat{x}_T + H\Phi^{-N+1}(\Xi_u U_{T-1,N-1} + \Xi_y Y_{T-1,N-1})),\end{aligned}\tag{5.24a}$$

$$\hat{y}_{T+1} = H\hat{x}_{T+1},\tag{5.24b}$$

where

$$z_k = w_k + H\bar{x}_k,\tag{5.25a}$$

$$w_k = \left(\int_{-\infty}^{\infty} \frac{(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \frac{y_k - H\bar{x}_k}{\sigma^2\nu + (y_k - H\bar{x}_k)^2}, \quad k = T - N + 1 \text{ or } T + 1,\tag{5.25b}$$

$$L_N = \Phi^{N-1} P_N (H\Phi^{N-1})',$$

$$\tilde{L}_N = \Phi^{N-1} P_N (H\Phi^{-1})',$$

$$P_N = \left(\sum_{i=1}^N (H\Phi^{i-1})' H\Phi^{i-1} \right)^{-1},$$

$$\bar{x}_{T+1} = \Phi\bar{x}_T + \Gamma u_T + \Omega y_T.$$

Proof. See Appendix 8. □

The recursive MHE-TD is a computationally efficient Kalman filter like recursive algorithm parameterized by the shape parameter of t -distribution, i.e., ν . As

shown in Remark 5.3 the recursive MHE-TD reduces to the recursive solution of the standard Moving Widow Least-Squares Estimation by simply setting $\nu = \infty$.

REMARK 5.3 *When the noise e_k in ARMAX process (2.7) is Gaussian (i.e., $\nu = \infty$) the recursive MHE-TD reduces to the following recursive solution of the standard Moving Window Least-Squares Estimation*

$$\begin{aligned}\hat{x}_{T+1} = & \Phi \hat{x}_T + \Gamma u_T + \Omega y_T + L_N (y_{T+1} - H(\Phi \hat{x}_T + \Gamma u_T + \Omega y_T)), \\ & - \tilde{L}_N (y_{T-N+1} - H\Phi^{-N+1} \hat{x}_T + H\Phi^{-N+1} (\Xi_u U_{T-1,N-1} + \Xi_y Y_{T-1,N-1}))\end{aligned}\quad (5.26a)$$

$$\hat{y}_{T+1} = H \hat{x}_{T+1}. \quad (5.26b)$$

Comparing (5.24a) and (5.26a) our recursive MHE-TD has the same structure as the standard Moving Window Least-Squares except a nonlinear transformation from measurement, i.e., (5.25b). When $\Omega = 0$, $\Gamma = 0$, Equation (5.26a) reduces to (10) in [72].

The MHE uses only a finite number of past measurement samples; the oldest measurement is discarded as a new sample becomes available. The ARMAX filter [53] uses the so-called growing memory strategy where the past measurement is not discarded as new measurement becomes available. Therefore, the ARMAX filter can be obtained if the last term of (5.24a) is discarded and L_N and P_N become

$$L_T = \Phi^{T-1} P_T (H \Phi^{T-1})', \quad P_T = \left(\sum_{i=1}^T (H \Phi^{i-1})' H \Phi^{i-1} \right)^{-1}$$

5.4 Statistical Interpretation

In this section we analyze the statistical properties of the proposed estimator. Let the actual noise to be associated with pdf $g(e)$ which is different from $f(e)$ used in

the design of MHE-TD. It is well known that the mean and variance of an estimator can be analysed by its IF [34, 47, 81]. The statistical properties of IF approximation of MHE-TD is derived by using IF as follows.

THEOREM 5.2 STATISTICAL PROPERTIES OF IF APPROXIMATION OF MHE-TD

Consider the estimator (5.19). Assume that the data are generated from the AR-MAX process (2.7) where the pdf of e_k is $g(e)$ which is assumed to be t distribution noise. The expectation and variance of the estimation are given by

$$E(\hat{x}_T) = \sum_{k=1}^{T-1} (\Phi + \Omega)^{T-k-1} \Gamma u_k, \quad (5.27a)$$

$$\begin{aligned} Var(\hat{x}_T) = & \frac{\rho_1}{\rho_4^2} M_N + 2 \sum_{i=1}^{N-1} \frac{\rho_2}{\rho_4} M_N (H \Phi^{i-N})' \left((\Phi + \Omega H)^{N-i-1} \Omega \right)' \\ & + \sum_{k=1}^{T-1} \rho_3 (\Phi + \Omega H)^{T-k-1} \Omega \left((\Phi + \Omega H)^{T-k-1} \Omega \right)', \end{aligned} \quad (5.27b)$$

where

$$\rho_1 = \int_{-\infty}^{\infty} \left(\frac{(\nu+1)e}{\sigma^2\nu + e^2} \right)^2 g(e) de, \quad \rho_2 = \int_{-\infty}^{\infty} \frac{(\nu+1)e^2}{\sigma^2\nu + e^2} g(e) de, \quad (5.28)$$

$$\rho_3 = \int_{-\infty}^{\infty} e^2 g(e) de, \quad \rho_4 = \int_{-\infty}^{\infty} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} g(e) de, \quad (5.29)$$

$$M_N = \left(\sum_{i=1}^N (H \Phi^{i-N})' H \Phi^{i-N} \right)^{-1}.$$

With $\hat{y}_T = H \hat{x}_T$,

$$E(\hat{y}_T) = H \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k, \quad (5.30a)$$

$$\begin{aligned} Var(\hat{y}_T) = & \frac{\rho_1}{\rho_4^2} H M_N H' + 2 \sum_{i=1}^{N-1} \frac{\rho_2}{\rho_4} H M_N (H \Phi^{i-N})' H (\Phi + \Omega H)^{N-i-1} \Omega \\ & + \sum_{k=1}^{T-1} \rho_3 \left(H (\Phi + \Omega H)^{T-k-1} \Omega \right)^2. \end{aligned} \quad (5.30b)$$

Proof. See Appendix 9. □

COROLLARY 5.2 *The expectation of Moving Window Least-Squares Estimation is given by*

$$E(\hat{x}_T) = \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k$$

and its variance is

$$\begin{aligned} Var(\hat{x}_T) = & \rho_3 \left(M_N + 2 \sum_{i=1}^{N-1} M_N (H \Phi^{i-N})' \left((\Phi + \Omega H)^{N-i-1} \Omega \right)' \right. \\ & \left. + \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Omega \left((\Phi + \Omega H)^{T-k-1} \Omega \right)' \right), \end{aligned} \quad (5.31)$$

where

$$\rho_3 = \int_{-\infty}^{\infty} e^2 g(e) d(e).$$

With $\hat{y}_T = H \hat{x}_T$,

$$\begin{aligned} E(\hat{y}_T) &= H \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k \\ Var(\hat{y}_T) &= \rho_3 \left(H M_N H' + 2 \sum_{i=1}^{N-1} H M_N (H \Phi^{i-N})' H (\Phi + \Omega H)^{N-i-1} \Omega \right. \\ & \quad \left. + \sum_{k=1}^{T-1} \left(H (\Phi + \Omega H)^{T-k-1} \Omega \right)^2 \right). \end{aligned} \quad (5.32)$$

Proof. Recall that ρ_1, ρ_2, ρ_3 and ρ_4 in (5.27b) are functions of ν as shown in (5.28) and (5.29). By setting $\nu = \infty$ in (5.27b). □

REMARK 5.4 *Corollary 5.2 derives the statistical property of Moving Window Least-Squares Estimation for ARMAX model. When $\Omega = 0$, Corollary 5.2 reduces to the well known result for statistical property of least-squares estimation considered in Theorem 13.2 in [10].*

The assumption that $g(e)$ is the same at each sampling instant k is commonly made. Here we extend to the case where $g(e)$ could be different at different k denoted by $g_k(e)$. This case is useful for analysis when outlier happens (see Simulation 3).

COROLLARY 5.3 PROPERTY OF IF APPROXIMATION OF MHE-TD WITH OUTLIER *The expectation of the estimator (5.19) given an outlier at k_1 is*

$$E(\hat{y}_T) = \begin{cases} HE(\bar{x}_T) + \frac{1}{\rho_4} HM_N \frac{(H\Phi^{k_1-T})' (\nu + 1)e_{k_1}}{\sigma^2\nu + e_{k_1}^2} & \text{if } k_1 \leq T \leq k_1 + N - 1, \\ HE(\bar{x}_T) & \text{if } T < k_1 \text{ or } T > k_1 + N - 1, \end{cases} \quad (5.33)$$

where

$$E(\bar{x}_T) = \begin{cases} \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k & \text{if } T \leq k_1, \\ \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k + (\Phi + \Omega H)^{T-k_1-1} \Omega e_{k_1} & T > k_1. \end{cases} \quad (5.34)$$

Proof. See Appendix 10. □

REMARK 5.5 *The expectation of Moving Window Least-Squares Estimation with outlier e_{k_1} can be obtained by setting $\nu = \infty$ in (5.33)*

$$E(\hat{y}_T) = \begin{cases} HE(\bar{x}_T) + HM_N (H\Phi^{k_1-T})' e_{k_1} & \text{if } k_1 \leq T \leq k_1 + N - 1, \\ HE(\bar{x}_T) & \text{if } T < k_1 \text{ or } T > k_1 + N - 1, \end{cases} \quad (5.35)$$

where $E(\bar{x}_T)$ is given in (5.34).

5.5 Simulations

Three simulations are given to illustrate the performance of IF approximation of MHE-TD. For easy reference, the parameters of the ARMAX process, IF approxima-

tion of MHE-TD and ARMAX filter are summarized in Table 5.1. The parameters Φ , H , Γ and Ω depend on A , B and C of the ARMAX process. Notice that the distribution $f(e)$ used for the design of estimator needs not be the same as $g(e)$, the distribution of the actual noise shown in the last row of Table 5.1.

Table 5.1: Parameters of the ARMAX process, the IF approximation of MHE-TD and ARMAX filter in the Simulations.

Simulation	1	2	3
Figure Number	-	5.4	5.5
Table Number	5.2	5.3	-
ARMAX Model	A , $1 + a_1q^{-1} + a_2q^{-2}$ $a_1 = -1.4, a_2 = 0.48$	$1 + aq^{-1}$, $a = -0.9$	$1 + aq^{-1}$ $a = -0.6$
	B bq^{-1} $b = 0.1$	bq^{-1} , $b = 0.1$	bq^{-1} $b = 0.1$
	C $1 + c_1q^{-1} + c_2q^{-2}$, $c_1 = -0.8, c_2 = 0.12$	$1 + cq^{-1}$, $c = -0.8$	$1 + cq^{-1}$ $c = -0.5$
	$f(e)$ $t_3(0, 0.1)$	$t_3(0, 0.1)$	$t_3(0, 0.1)$
MHE-TD for ARMAX Process	Φ_a $\begin{bmatrix} -a_1 & 1 \\ -a_2 & 0 \end{bmatrix}$	$-a$	$-a$
	H $\begin{bmatrix} 1 & 0 \end{bmatrix}$	1	1
	Γ $\begin{bmatrix} b \\ 0 \end{bmatrix}$	b	b
	Ω $\begin{bmatrix} c_1 - a_1 \\ c_2 - a_2 \end{bmatrix}$	$c - a$	$c - a$
	ν 3	3	3
	σ 0.1	0.1	0.1
	N 3, 6, 9	3, 6, 9	3
	Actual Noise $g(e)$	$f(e)$	$\begin{cases} \Delta_{e_{12}}, & \text{if } k = 12, e_{12} = -2, \\ f(e) & \text{if } k \neq 12. \end{cases}$

5.5.1 Simulation 1: Variance

Table 5.2 shows the variances from 1000 simulation runs each for Equation (5.19), ARMAX filter [53], Kalman filter and Moving Window Least-Squares Estimation (Equation (5.26)). In these simulations, the measurement noise was t_3 with $\nu = 3$ and $\sigma = 0.1$. IF approximation of MHE-TD and ARMAX filter is designed with

noise pdf which matches exactly with the measurement noise. We perform different simulations for different horizon length of N .

Table 5.2 contains the variances obtained from simulation runs and theoretical calculations. From Table 5.2 we can find that the match between the theoretical results, i.e., Equations (5.30b)¹ and (5.32), and the simulation results. Thus Theorem 5.2 can be used to quantify how the variance of estimate changes when N or noise parameters change. For this particular example we can see from Table 5.2 that when N increases the variances of both IF approximation of MHE-TD and Moving Window Least-Squares Estimation decrease. In other words, the moving window length of N can be used as a tuning parameter for IF approximation of MHE-TD. In addition, for the same N the variance of estimate obtained by Moving Window Least-Squares Estimation is bigger than IF approximation of MHE-TD. Table 2 confirms the fact that the estimators with growing memory, i.e., ARMAX filter and Kalman filter, gives smaller estimate variance in steady state comparing with the the estimators with fixed memory, i.e., IF approximation of MHE-TD and Moving Window Least-Squares estimation. From Table 5.2 we can also find that IF approximation of MHE-TD becomes ARMAX filter if the horizon N grows with T , i.e., $N = T$. This is because IF approximation of MHE-TD is the moving horizon version of ARMAX filter.

In the following the variances of estimation of IF approximation of MHE-TD and Moving Window Least-Squares Estimation at $T = 20$ versus N with $p = 2$, $q = 1.5$, $\sigma = 0.1\sqrt{2}$ and the noise shape parameter q with $N = 5$ and $p = 2$, $\sigma = 0.1\sqrt{2}$ are shown in Figures 5.2 and 5.3, respectively.

As shown in Figure 5.2 the variance of estimation of IF approximation of MHE-TD and Moving Window Least-Squares Estimation decrease as N increasing, and the variance of estimation of IF approximation of MHE-TD is always smaller than

¹by substituting $\nu = 3$, $\sigma = 0.1$

Table 5.2: Variance of $\hat{y}_T$ in Simulation 1.

T			3	6	9	20	40	60
IF approx- imation of MHE-TD	theoretical calculation from (5.30)	$N = 3$	0.0312	0.0403	0.0426	0.0435	0.0435	0.0435
		$N = 6$	-	0.0298	0.0322	0.0330	0.0330	0.0330
		$N = 9$	-	-	0.0295	0.0303	0.0303	0.0303
	Simulation results	$N = 3$	0.0312	0.0403	0.0426	0.0435	0.0435	0.0435
		$N = 6$	-	0.0298	0.0322	0.0330	0.0330	0.0330
		$N = 9$	-	-	0.0295	0.0303	0.0303	0.0303
Moving Window Least- Squares estimation	theoretical calculation from (5.32)	$N = 3$	0.0448	0.0538	0.0562	0.0570	0.0570	0.0570
		$N = 6$	-	0.0329	0.0353	0.0361	0.0361	0.0361
		$N = 9$	-	-	0.0299	0.0307	0.0307	0.0307
	Simulation results	$N = 3$	0.0448	0.0538	0.0562	0.0570	0.0570	0.0570
		$N = 6$	-	0.0329	0.0353	0.0361	0.0361	0.0361
		$N = 9$	-	-	0.0299	0.0307	0.0307	0.0307
ARMAX filter	Simulation result	-	0.0312	0.0298	0.0295	0.0293	0.0293	0.0293
Kalman filter	Simulation result	-	0.0448	0.0329	0.0299	0.0296	0.0293	0.0293

Figure 5.2: Variance of estimation versus N . Figure 5.3: Variance of estimation versus noise shape parameter q .

Moving Window Least-Squares Estimation during N varying. The figure also verifies that when $N \rightarrow \infty$, IF approximation of MHE-TD is equivalent to Moving Window Least-Squares Estimation.

In Figure 5.3 the variance of estimation of IF approximation of MHE-TD and Moving Window Least-Squares Estimation decrease as q increasing. And the variance of estimation of IF approximation of MHE-TD is always smaller than Moving Window Least-Squares Estimation during q varying. The figure also verifies that when the noise is Gaussian, i.e., $\nu = pq \rightarrow \infty$, IF approximation of MHE-TD is equivalent to Moving Window Least-Squares Estimation.

5.5.2 Simulation 2: Model Mismatch

Figure 5.4 shows compares the performance of IF approximation of MHE-TD with, Moving Window Least-Squares estimation with, ARMAX filter and Kalman filter. In the simulation the measurement noise is t_3 noise with $\nu = 3$ and $\sigma = 0.1$. Suppose the plant parameters change in the following way:

$$a = \begin{cases} -0.9 & \text{if } k \leq 25 \text{ or } k \geq 29, \\ -0.6 & \text{if } 25 < k < 29, \end{cases}, \quad c = \begin{cases} -0.8 & \text{if } k \leq 25 \text{ or } k \geq 29, \\ -0.3 & \text{if } 25 < k < 29. \end{cases}$$

The plant model parameters used in IF approximation of MHE-TD, Moving Window Least-Squares estimation, ARMAX filter and Kalman filter are $a = -0.9$, $c = -0.8$. IF approximation of MHE-TD and ARMAX filter [53] are designed with $\nu = 3$ and $\sigma = 0.1$.

Figure 5.4 shows that the estimation error of IF approximation of MHE-TD converges faster than ARMAX filter and Kalman filter but less than Moving Window Least-Squares estimation. Thus IF approximation of MHE-TD is more robust to model mismatch than ARMAX filter and Kalman filter but less robust than Moving Window Least-Squares estimation.

Table 5.3 shows the variances from 1000 simulation runs each for Equation (5.19), ARMAX filter [53], Kalman filter and Moving Window Least-Squares Estimation (Equation (5.26)). We perform different simulations for different horizon length of N .

In Table 5.3, for the same N , IF approximation of MHE-TD gives smaller estimate variance in steady state than Moving Window Least-Squares estimation but bigger estimate variance than ARMAX filter and Kalman filter. This is useful for the scenario where both model mismatch happens and a specific estimate accuracy needs. For example according to Theorem 5.2 smaller N which means less memory is needed can be determined to meet the estimate accuracy while the moving horizon structure makes the proposed estimator to be robust to model mismatch. In addition, the parameter N can be used to compromise the estimation variance and recovery time of the IF approximation of MHE-TD after uncertainty happens. For example, comparing Figure 5.4(a) where $N = 3$ and 5.4(c) where $N = 9$, the IF approximation of MHE-TD with $N = 3$ recovers faster than the IF approximation of MHE-TD with $N = 9$. However, the estimation variance with $N = 3$ is bigger than $N = 9$.

5.5.3 Simulation 3: Outlier

Figure 5.5(a) and 5.5(b) show 1000 simulation runs each for IF approximation of MHE-TD (Equation (5.19)) and Moving Window Least-Squares Estimation (Equation (5.26)), respectively. In these simulations, the measurement noise was t_3 with $\nu = 3$ and $\sigma = 0.1$ except for an outlier of $e_{k_1} = -2$ at $k_1 = 12$. Hence $g(e)$ is given as

$$g(e) = \begin{cases} \Delta_{e_{k_1}} & \text{if } k = k_1, \\ f(e) & \text{if } k \neq k_1. \end{cases}$$

Figure 5.4: Estimation error of IF approximation of MHE-TD, Moving Window Least-Squares estimation, ARMAX filter and Kalman filter when model mismatch happened at $25 < T < 29$: (a) $N=3$; (b) $N=6$; (c) $N=9$.

Table 5.3: Variance of $\hat{y}_T$ in Simulation 2.

T			3	6	9	20	40	60
IF approx- imation of MHE-TD	theoretical calculation from (5.30)	$N = 3$	0.0051	0.0056	0.0058	0.0061	0.0062	0.0062
		$N = 6$	-	0.0026	0.0029	0.0031	0.0032	0.0032
		$N = 9$	-	-	0.0019	0.0021	0.0022	0.0022
	Simulation results	$N = 3$	0.0051	0.0056	0.0058	0.0061	0.0062	0.0062
		$N = 6$	-	0.0026	0.0029	0.0031	0.0032	0.0032
		$N = 9$	-	-	0.0019	0.0021	0.0022	0.0022
Moving Window Least- Squares estimation	theoretical calculation from (5.32)	$N = 3$	0.0097	0.0102	0.0104	0.0106	0.0107	0.0107
		$N = 6$	-	0.0042	0.0045	0.0047	0.0048	0.0048
		$N = 9$	-	-	0.0025	0.0028	0.0029	0.0029
	Simulation results	$N = 3$	0.0097	0.0102	0.0104	0.0106	0.0107	0.0107
		$N = 6$	-	0.0042	0.0045	0.0047	0.0048	0.0048
		$N = 9$	-	-	0.0025	0.0028	0.0029	0.0029
ARMAX filter	Simulation result	-	0.0051	0.0026	0.0019	0.0016	0.0015	0.0015
Kalman filter	Simulation result	-	0.0097	0.0042	0.0025	0.0016	0.0015	0.0015

IF approximation of MHE-TD is designed with $\nu = 3$ and $\sigma = 0.1$ which is the same as measurement noise. Figure 5.5 shows the results of the 1000 simulation runs, and the white curve is the mean value of the 1000 runs, which coincides with the theoretical formula shown below.

Substituting the parameters in Simulation 3 to equations (5.33) and (5.35), and simplifying give the mean of IF approximation of MHE-TD

$$\begin{aligned}
& E(\hat{y}_T) \\
= & \begin{cases} HE(\bar{x}_T) + \frac{(-c)^{k_1-T}(c^{2(N-1)} - c^{2N})}{\rho_4(1 - c^{2N})} \left(\frac{4e_{k_1}}{0.03 + e_{k_1}^2} \right) & \text{if } k_1 \leq T \leq k_1 + N - 1, \\ HE(\bar{x}_T) & \text{if } T < k_1 \text{ or } T > k_1 + N - 1 \end{cases}
\end{aligned} \tag{5.36}$$

Figure 5.5: Simulation results of Simulation 3: (a) $\hat{y}_T$ of IF approximation of MHE-TD; (b) $\hat{y}_T$ of Moving Window Least-Squares Estimation.

and the mean of Moving Window Least-Squares Estimation

$$E(\hat{y}_T) = \begin{cases} HE(\bar{x}_T) + \frac{(-c)^{k_1-T}(c^{2(N-1)} - c^{2N})}{1 - c^{2N}} e_{k_1} & \text{if } k_1 \leq T \leq k_1 + N - 1, \\ HE(\bar{x}_T) & \text{if } T < k_1 \text{ or } T > k_1 + N - 1, \end{cases} \quad (5.37)$$

where

$$\rho_4 = \int_{-\infty}^{\infty} \frac{0.12 - 4e^2}{(0.03 + e^2)} g(e) de = \frac{200}{3}.$$

5.6 Experiments

5.6.1 Experiment 1: Liquid Level Estimation

In this experiment the liquid level of a Coupled-Tank Control Apparatus is estimated by the IF approximation of MHE-TD and Moving Window Least-Squares Estimation.

Experiment Description

Figure 5.6: Experiment apparatus: Coupled-Tank Control Apparatus.

The Coupled-Tank Control Apparatus is shown in Figure 5.6 where the liquid level of Tank 1 is to be estimated. The process model of Tank 1 is identified as

$$y_k = \frac{0.066q^{-1}}{1 - 0.99q^{-1}}u_k + e_k \quad (5.38)$$

where y_k is the measurement of the liquid level in the tank and e_k is the measurement noise. With setting the control input voltage $u_k = 2 \text{ V}$, 1000 measurements are shown in the Figure 5.7.

Figure 5.7: Measurements of Experiment y_k .

Fitting of the distribution of noise

The maximum likelihood criteria is used to identify the parameters of t -distribution. The noise e_k is obtained by subtracting the mean of y_k from measurement y_k . In this identification we find ν and σ of t -distribution probability density function $f(e)$ of Equation (5.1) by maximizing the following objective function

$$J = \sum_{i=1}^{1000} \ln \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\sigma\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{e_i^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}}$$

This gives $\nu = 4.2054$ and $\sigma = 0.0792$. The resultant t -distribution is superimposed on the distribution in Figure 5.8. Using maximum likelihood criteria a Gaussian distribution is also obtained to fit to the histogram of experimental noise e_k . Figure 5.8 shows t -distribution gives better fit.

Figure 5.8: Noise distribution and fitting.

Estimation results of y_k

With the control input voltage $u_k = 2$ V, the liquid level y_k for 10 samples is estimated by the IF approximation of MHE-TD and Moving Window Least-Squares Estimation with $N = 5$. We run 100 simulation runs and the results are shown in Figures 5.9 and 5.10.

The variance of estimation results at $T = 5, 8, 10$ in Figures 5.9 and 5.10 are tableted in Table 5.4 which show that the variance from the IF approximation of

Figure 5.9: Estimation result: Estimation of Liquid Level by using Moving Window Least-Squares Estimation.

Figure 5.10: Estimation result: Estimation of Liquid Level by using the IF approximation of MHE-TD.

MHE-TD is smaller than Moving Window Least-Squares Estimation.

Consider the plant in the experiment of liquid level estimation. Suppose the measurement noise is t -distribution noise with $p = 2$, $q = 1.5$, $\sigma = 0.2\sqrt{2}$. Figures 5.11 and 5.12 show 1000 simulation runs each for IF approximation of MHE-TD and Moving Window Least-Squares Estimation, respectively. From Figures 5.11 and 5.12 the same conclusion as the one obtained in the experiment is obtained, i.e., the estimation variance of IF approximation of MHE-TD is smaller than the estimation variance of Moving Window Least-Squares Estimation.

Table 5.4: Variance of $\hat{y}_T$ in Figures 5.9 and 5.10

T	IF approximation of MHE-TD	Analysis	Moving Window Least-Squares Estimation	Analysis
5	0.0015	0.0017	0.0024	0.0023
8	0.0017	0.0017	0.0031	0.0023
10	0.0018	0.0017	0.0023	0.0023

Figure 5.11: $\hat{y}_k$ of IF approximation of MHE-TD.

Figure 5.12: $\hat{y}_k$ of Moving Window Least-Squares Estimation.

5.6.2 Experiment 2: Indoor environment localization

In this experiment the proposed IF approximation of MHE-TD is employed in the indoor environment localization.

Experiment Description

Figure 5.13: Indoor environment localization test-bed.

Figure 5.14: The histogram of noise e_k .

Figure 5.13 shows the indoor environment test-bed where six signal receivers (five-pointed star) are located at known positions. A person taking a transmitter walks from point $(0, 10)$ to point $(0, 0)$. The measurement sampling time is $h = 1$ sec. During the process of walking, three best sensors (with the strongest signal strength) are selected to do the trilateration. The histogram of noise e_k is shown in Figure 5.14. The results of the trilateration are shown in Figure 5.15.

Figure 5.15: The results of the trilateration.

Estimation results

A single-input single-output double integrator ARMAX model used in the IF approximate of MHE-TD and Moving Window Least-Squares Estimation is (2.7) with

$$\begin{aligned} A(q^{-1}) &= (1 - q^{-1})^2, \quad B(q^{-1}) = 0, \\ C(q^{-1}) &= 1 - 1.8q^{-1} + 0.81q^{-2}. \end{aligned} \quad (5.39)$$

Expressing Equation (2.7) with (5.39) in the state-space forms gives the constant velocity state-space model

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} 0.2 \\ \frac{0.01}{h} \end{bmatrix} e_k \quad (5.40)$$

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + e_k \quad (5.41)$$

Substituting Equation (5.41) and $h = 1$ to (5.40) gives

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 0.8 & 1 \\ -0.01 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0.01 \end{bmatrix} y_k \quad (5.42)$$

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + e_k \quad (5.43)$$

where y_k is the results of the trilateration. After fitting of the distribution of noise e_k as did in the Experiment 1, the noise e_k is characterized as $t_{2,3781}$ noise with $\sigma = 0.1635$.

We compare the estimation results of IF approximation of MHE-TD and Moving Window Least-Squares Estimation in Figure 5.16 which shows that the proposed IF approximation of MHE-TD is less sensitive to the outlier than Moving Window Least-Squares Estimation.

Consider the plant in the experiment of the indoor environment localization. Suppose an outlier happened at $k = 30$ and the initial state $x_0 = \begin{bmatrix} 10 & 0.05 \end{bmatrix}'$. Figure 5.17 shows estimation results of IF approximation of MHE-TD and Moving Window Least-Squares Estimation. From Figure 5.17 the same conclusion as the one obtained in the experiment is obtained, i.e., the performance of IF approximation

Figure 5.16: Estimation results of IF approximation of MHE-TD and Moving Window Least-Squares Estimation.

of MHE-TD is less affected by the outlier than Moving Window Least-Squares Estimation.

Figure 5.17: $\hat{y}_k$ of IF approximation of MHE-TD and Moving Window Least Squares Estimation.

5.7 Conclusion

The IF is employed to give an approximate solution to moving horizon MLE for ARMAX process with t -distribution noise, i.e., IF approximation of MHE-TD. The approximate solution can be formulated as the recursive MHE-TD algorithm which

makes it suitable for on-line and real-time implementation. We also used the IF to analyze the output of the proposed estimator designed with the t -distribution noise model instead of the usual Gaussian noise model. Formulas derived are useful in determining the variance of the estimates and the impact of outliers. As expected the proposed estimator reduces to Kalman filter when the noise is modeled by the Gaussian distribution and the horizon grows with time instant. The simulations show the advantages of the proposed estimator over Moving Horizon Least-Squares Estimation in the sense of estimation variance and the advantage over Kalman filter and ARMAX filter when model mismatch happens. The two experiments show that the proposed estimator gives smaller estimation variance and is more robust to outlier than Moving Window Least-Square estimation. IF approximation of MHE-TD gives a compromise estimate result which has smaller variance than Moving Window Least-Squares estimation and is more robust to model mismatch than Kalman filter and ARMAX filter. This is useful for the scenario where both model mismatch happens and a specific estimate accuracy needs.

Chapter 6

Conclusion and Future Work

In this thesis, We showed the relationship between the trajectory advisor for MPC with tracking error bound and the measurement checker for constrained MHE. The trajectory advisor can supply a feasible reference signal to MPC when the original reference signal makes MPC infeasible. The measurement checker for constrained MHE is obtained by symmetry. The measurement checker can determine the admissibility of the measurements before the constrained MHE is solved. In addition, a novel MHE for ARMAX process with t -distribution noise by using IF approach is proposed. Some possible future work are suggested below:

1. For the MPC with tracking error bound, so far we only investigated the behavior of the MPC and its accompanied trajectory advisor, assuming that the initial feasible set has been obtained. The computational complexity of obtaining the initial feasible set has not been discussed. The computation of the initial feasible set is done by projection of the sets. In literature approximate computation methods of the (robust) initial feasible set have been investigated in [14, 57, 60, 63, 96]. These techniques may be applicable in obtaining the initial feasible set for our proposed trajectory advisor.
2. In this thesis we adopted integrator augmentation to delete steady state error.

The integrator augmentation is only a simple method. The internal principle approach is a more general approach. In the future work we will investigate the application of the internal principle approach on the MPC with trajectory advisor.

3. In order to maintain the symmetry between MPC with MHE, we only use the receding horizon strategy to update the reference which results that the degree of freedom to ensure the feasibility of the MPC is to adjust y_N^r in the trajectory advisor. As far as we know, the other reference updating scheme where the entire reference sequence over the prediction horizon is updated has not been studied. In the future we would consider this reference updating scheme. Then the degree of freedom to ensure the feasibility is to adjust $y_0^r, y_1^r, \dots, y_N^r$ which means more degree of freedom. In [15] the author also hint what we suggest but did not illustrate the performance. This more general version of trajectory advisor may give better tracking performance of MPC and we will investigate it in future work.

4. For the proposed MHE, the Influence Function is used to formulate an approximate solution to the moving horizon MLE for ARMAX process with t -distribution noise. However, we did not discuss how good the approximation is. Consider the case when $\Phi_a = 0.8$, $\Gamma = 0$, $\Omega = 0.2$, $H = 1$, and $T = 3$, $\nu = 3$, $\sigma = 1/\sqrt{2}$, $y_1 = 1$, $y_2 = 2$ and $y_3 = 1$ in the ARMAX process (5.2). In Figure 6.1(a), the solid line shows the MLE for ARMAX process with t_3 distribution noise as a function of $h \in [0, 1]$ and $\kappa(1) = 1.45$ which is the solution of MLE (5.5), i.e., $\hat{x}_1 = 1.45$.

Figure 6.1: (a) MLE estimate as a function of h (solid line) and the tangent at $h = 0$ (dashed line) for t_3 noise; (b) Second order derivative of $\kappa(h)$ wrt. h .

Substituting $T = 3$, $\nu = 3$, $\sigma = 1/\sqrt{2}$, $y_1 = 1$, $y_2 = 2$, $y_3 = 1$ to (5.16) gives

$$x_{\text{IF}} = \text{IF}(e) = 1.92,$$

where $\text{IF}(e)$ gives the gradient of the tangent line of the solid line at $h = 0$, i.e., the red dashed line, in Figure 6.1(a). According to [64] the remainder of the first order Taylor series expansion, i.e., (5.12) is given by

$$R_1(1) = \frac{1}{2!} \frac{d^2\kappa(h)}{dh^2} \Big|_{h=h_L} (1-0)^2 \quad \text{for some } h_L \in [0, 1],$$

where $\frac{d^2\kappa(h)}{dh^2}$ is shown in Figure 6.1(b) from which we can determine

$$-1.07 \leq R_1(1) \leq 0.68.$$

which means the worst approximation error is -1.07 which is a $|-1.07/1.45| = 74\%$ error.

In general the approximation error depends on the particular set of measurements and the number of the measurements. If the number of the measurements goes to infinity, the IF approximation will get the true mean of the data which means no approximation. In order to analysis the approximation error, in statistics, the Von Mises expansions [81] of the statistical functional $\kappa(h)$

will be used in the future.

5. In this thesis, we did not analyse the convergence properties of the proposed first order Taylor series approximation used in our MHE. This is because convergence analysis is not needed for the current work. Convergence means whether the Taylor series converges to the true value of the function as more terms are added. One topic for future research is to analyse the convergence properties of our proposed approximation and to obtain the bounds of the approximation error.

In the related works [104, 105], the optimization problem in Maximum Likelihood Estimation is solved numerically which needs much computational effort. In our proposed method, the solution to the optimization problem is approximated by the first order Taylor series. The approximate solution can be written as a Kalman filter-like recursive algorithm which is more computationally efficient than the numerical methods. In the future, more studies could be carried out to compare the performance of our method with the numerical methods in the related works.

6. We have carried out preliminary investigation and observation on the effect of S/H time on the size of initial feasible set of standard MPC regulation problem. And we only consider the case that the S/H time h can be dividable by the continuous time prediction horizon T . The general case can be found in [11]. The analysis of the relationship between the size of the initial feasible set and the S/H time is not addressed. In [83] the effect of S/H time on the rate of the optimal discrete time cost converges to optimal continuous time cost has been investigated. This may be helpful for us to analysis the effect of S/H time on the size of initial feasible set in the future.

AUTHOR'S PUBLICATIONS

1. **Dexiang Zhou**, K.V. LING, W.K. Ho and Jan M. Maciejowski, "Moving Horizon Estimation for ARMAX process with t-Distribution Noise". First round revision submitted to *Automatica*, 2016.
2. **Dexiang Zhou**, K.V. LING, A Trajectory Advisor for Model Predictive Control with Tracking Error Bound and a Measurement Checker for Moving Horizon Estimation. In preparation to be submitted to *International Journal of Control*, 2016.
3. **Dexiang Zhou** and K.V. Ling, "The Effect of Sample/Hold Time on Initial Feasible Set in Model Predictive Control Design" *2013 Australian Control Conference*, (Perth, Western Australia).
4. **Dexiang Zhou**, K.V. LING and E.K. Poh, "Constrained Moving Horizon Estimation for Linear and Nonlinear Systems" *3rd Workshop SONDRALaLonde 2013*, (La Londe les Maures, France).

Not reported in this thesis:

5. Tri Tran, **Dexiang Zhou**, K-V. Ling, Jan M. Maciejowski, Multiplexed Model Predictive Control of Interconnected Systems, *IEEE Conference on Decision and Control*, 2015.

6. Tengpeng Chen, **Dexiang Zhou**, Tri Tran, Catharine Kastner, K-V. Ling, K-J. Tseng, and Jan M. Maciejowski, “Distributed moving horizon estimation for power systems,” IEEE PES GENERAL MEETING, Denver, CO, USA, 26–30 July, 2015.
7. Tengpeng Chen, **Dexiang Zhou**, Tri Tran, W-K. Ho, and K-V. Ling, “Iterated moving horizon estimation for power systems,” Proceedings of IEEE International Conference on Control, Automation, and Information Science (ICCAIS 14) Gwangju, South Korea, 2-5 December, 2014.

Bibliography

- [1] Gabriel Agamennoni, Juan I Nieto, and Eduardo Mario Nebot. An outlier-robust Kalman filter. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1551–1558, 2011.
- [2] Gabriel Agamennoni, Juan I Nieto, and Eduardo M Nebot. Approximate inference in state-space models with heavy-tailed noise. *Signal Processing, IEEE Transactions on*, 60(10):5024–5037, 2012.
- [3] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44(7):1753–1765, 2008.
- [4] Frank Allgöwer and Alex Zheng. *Nonlinear model predictive control*. Birkhäuser, 2000.
- [5] Ignacio Alvarado, Daniel Limon, Teodoro Alamo, and Eduardo Camacho. On the design of robust tube-based MPC for tracking. In *World Congress*, volume 17, pages 15333–15338, 2008.
- [6] Rishi Amrit, James B Rawlings, and David Angeli. Economic optimization using model predictive control with a terminal cost. *Annual Reviews in Control*, 35(2):178–186, 2011.
- [7] David Angeli, Rishi Amrit, and James B Rawlings. On average performance

- and stability of economic model predictive control. *Automatic Control, IEEE Transactions on*, 57(7):1615–1626, 2012.
- [8] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188, 2002.
- [9] Karl J Åström and Björn Wittenmark. *Adaptive control*. Courier Dover Publications, 2013.
- [10] Karl Johan Åström and Björn Wittenmark. *Computer-controlled systems: theory and design*. Courier Dover Publications, 2011.
- [11] Vincent Bachtar, Eric C Kerrigan, William H Moase, and Chris Manzie. Smoothness properties of the mpc value function with respect to sampling time and prediction horizon. In *Control Conference (ASCC), 2015 10th Asian*, pages 1–6. IEEE, 2015.
- [12] Alberto Bemporad and Edoardo Mosca. Fulfilling hard constraints in uncertain linear systems by reference managing. *Automatica*, 34(4):451–461, 1998.
- [13] Alberto Bemporad, Alessandro Casavola, and Edoardo Mosca. Nonlinear control of constrained linear systems via predictive reference management. *Automatic Control, IEEE Transactions on*, 42(3):340–349, 1997.
- [14] H Benlaoukli and S Olaru. Model predictive control–numerical methods for the invariant sets approximation. In *Numerical Analysis and Its Applications*, pages 192–199. Springer, 2009.
- [15] Hichem Benlaoukli, Sorin Olaru, S Marinkovic, and P Boucher. Invariant set constructions for feasible reference tracking. In *Control and Automation, 2009. ICCA 2009. IEEE International Conference on*, pages 972–977, 2009.

- [16] Giulio Betti, Marcello Farina, and Riccardo Scattolini. A robust MPC algorithm for offset-free tracking of constant reference signals. *Automatic Control, IEEE Transactions on*, 58(9):2394–2400, 2013.
- [17] Robert R. Bitmead, Michel Gevers, and Vincent Wertz. *Adaptive optimal control: The thinking man's GPC*. Prentice Hall, 1991.
- [18] F Blanchin and Walter Ukovich. Linear programming approach to the control of discrete-time periodic systems with uncertain inputs. *Journal of Optimization Theory and Applications*, 78(3):523–539, 1993.
- [19] Franco Blanchini. Set invariance in control. *Automatica*, 35(11):1747–1767, 1999.
- [20] George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. *Time series analysis: forecasting and control*. John Wiley & Sons, 2013.
- [21] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [22] A Bruckstein and Thomas Kailath. Recursive limited memory filtering and scattering theory (corresp.). *Information Theory, IEEE Transactions on*, 31(3):440–443, 1985.
- [23] S. Di Cairano and F. Borrelli. Reference tracking with guaranteed error bound for constrained linear systems. *IEEE Transactions on Automatic Control*, 2015.
- [24] E.F Camacho and C. Bordons. *Model predictive control*. Springer Berlin, 1999.
- [25] T. Chen and B.A. Francis. *Optimal sampled-data control systems*, volume 124. Springer London, 1995.

-
- [26] Luigi Chisci and Giovanni Zappa. Dual mode predictive tracking of piecewise constant references for constrained linear systems. *International Journal of Control*, 76(1):61–72, 2003.
- [27] M. Diehl, H. Ferreau, and N. Haverbeke. Efficient numerical methods for nonlinear mpc and moving horizon estimation. *Nonlinear Model Predictive Control*, 384:391–417, 2009.
- [28] Moritz Diehl, Rishi Amrit, and James B Rawlings. A lyapunov function for economic optimizing model predictive control. *Automatic Control, IEEE Transactions on*, 56(3):703–707, 2011.
- [29] Matthew Ellis, Helen Durand, and Panagiotis D Christofides. A tutorial review of economic model predictive control methods. *Journal of Process Control*, 24(8):1156–1178, 2014.
- [30] Paola Falugi. Model predictive control for tracking randomly varying references. *International Journal of Control*, 88(4):745–753, 2015.
- [31] Paola Falugi and David Q Mayne. Tracking performance of model predictive control. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 2631–2636, 2012.
- [32] M. Farina, G. Ferrari-Trecate, C. Romani, and R. Scattolini. Moving horizon estimation for distributed nonlinear systems with application to cascade river reaches. *Journal of Process Control*, 21(5):767–774, 2011.
- [33] Timm Faulwasser and Rolf Findeisen. A model predictive control approach to trajectory tracking problems via time-varying level sets of lyapunov functions. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 3381–3386, 2011.

- [34] Luisa Turrin Fernholz. On multivariate higher order von mises expansions. *Metrika*, 53(2):123–140, 2001.
- [35] Antonio Ferramosca, Daniel Limón, Alejandro H González, Darci Odloak, and Eduardo F Camacho. MPC for tracking target sets. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 8020–8025, 2009.
- [36] Antonio Ferramosca, Daniel Limón, Ignacio Alvarado, Teodoro Alamo, F Castaño, and Eduardo F Camacho. Optimal MPC for tracking of constrained linear systems. *International Journal of Systems Science*, 42(8):1265–1276, 2011.
- [37] E.G. Gilbert and K.T. Tan. Linear systems with state and control constraints: The theory and application of maximal output admissible sets. *Automatic Control, IEEE Transactions on*, 36(9):1008–1020, 1991.
- [38] Elmer Gilbert and Ilya Kolmanovsky. Nonlinear tracking control in the presence of state and control constraints: a generalized reference governor. *Automatica*, 38(12):2063–2073, 2002.
- [39] Elmer G Gilbert, Ilya Kolmanovsky, and Kok Tin Tan. Nonlinear control of discrete-time linear systems with state and control constraints: A reference governor with global convergence properties. In *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, volume 1, pages 144–149, 1994.
- [40] Elmer G Gilbert, Ilya Kolmanovsky, and Kok Tin Tan. Discrete-time reference governors and the nonlinear control of systems with state and control constraints. *International Journal of Robust and Nonlinear Control*, 5(5):487–504, 1995.

-
- [41] Elmer Grant Gilbert and Ilya V Kolmanovsky. Fast reference governors for systems with state and control constraints and disturbance inputs. 1999.
- [42] G.C. Goodwin, J.A. De Doná, M.M. Seron, and X.W. Zhuo. Lagrangian duality between constrained estimation and control. *Automatica*, 41(6):935–944, 2005.
- [43] Graham C Goodwin, María M Seron, and José A De Doná. *Constrained control and estimation: an optimisation approach*. Springer, 2006.
- [44] P-O Gutman and Michael Cwikel. Admissible sets and feedback control for discrete-time linear dynamical systems with bounded controls and states. *Automatic Control, IEEE Transactions on*, 31(4):373–376, 1986.
- [45] Sohrab Haghghat, Stefano Di Cairano, Dmytro Konobrytskyi, and Scott Bortoff. Coordinated control of a dual-stage positioning system using constrained model predictive control. In *ASME 2014 Dynamic Systems and Control Conference*, pages V001T02A005–V001T02A005. American Society of Mechanical Engineers, 2014.
- [46] FR Hampel. *Contributions to the theory of robust statistics*. PhD thesis, Ph. D. thesis, University of California, Berkeley.(1974),” The influence curve and its role in robust estimation.” *Journal of the American Statistical Association*, 69, 383-394.(1978),” Optimally bounding the Gross-Error-Sensitivity and, 1968.
- [47] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust statistics: the approach based on influence functions*, volume 114. Wiley, 2011.
- [48] Frank R Hampel, Peter J Rousseeuw, and Elvezio Ronchetti. The change-of-variance curve and optimal redescending M-estimators. *Journal of the American Statistical Association*, 76(375):643–648, 1981.
-

- [49] E.L. Haseltine and J.B. Rawlings. Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Industrial & engineering chemistry research*, 44(8):2451–2460, 2005.
- [50] N. Haverbeke, M. Diehl, and B. De Moor. A structure exploiting interior-point method for moving horizon estimation. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 1273–1278, 2009.
- [51] Mohsen Heidarinejad, Jinfeng Liu, and Panagiotis D Christofides. Economic model predictive control of nonlinear process systems using lyapunov techniques. *AIChE Journal*, 58(3):855–870, 2012.
- [52] Martin Herceg, Michal Kvasnica, Colin N Jones, and Manfred Morari. Multi-parametric toolbox 3.0. In *Control Conference (ECC), 2013 European*, pages 502–510. IEEE, 2013.
- [53] Weng Khuen Ho, Keck Voon Ling, Hoang Dung Vu, and Xiaoqiong Wang. Filtering of the armax process with generalized t-distribution noise: The influence function approach. *Industrial & Engineering Chemistry Research*, 2014.
- [54] Tobias Gybel Hovgaard, Kristian Edlund, and JB Jorgensen. Economic mpc for power management in the smartgrid. In *21st European Symposium on Computer Aided Process Engineering*, volume 29, pages 1839–1843. Elsevier Science, 2011.
- [55] Peter J Huber et al. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.
- [56] Andrew H Jazwinski. Stochastic processes and filtering theory, 1970.

-
- [57] Colin Jones, Eric C Kerrigan, and Jan Maciejowski. Equality set projection: A new algorithm for the projection of polytopes in halfspace representation. Technical report, Cambridge University Engineering Dept, 2004.
- [58] Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. *Journal of Fluids Engineering*, 83(1):95–108, 1961.
- [59] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.
- [60] Eric C. Kerrigan. *Robust constraint satisfaction: Invariant sets and predictive control*. PhD thesis, Cambridge, 2000.
- [61] Kiminao Kogiso and Kenji Hirata. Reference governor for constrained systems with time-varying references. *Robotics and Autonomous Systems*, 57(3):289–295, 2009.
- [62] Samuel Kotz and Saralees Nadarajah. *Multivariate t-distributions and their applications*. Cambridge University Press, 2004.
- [63] Koustas I Kouramas, Sasa V Rakovic, Eric C Kerrigan, JC Allwright, and David Q Mayne. On the minimal robust positively invariant set for linear difference inclusions. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC’05. 44th IEEE Conference on*, pages 2296–2301, 2005.
- [64] Erwin Kreyszig. *Advanced engineering mathematics*. John Wiley & Sons, 1988.
- [65] P. Kühn, M. Diehl, T. Kraus, J.P. Schlöder, and H.G. Bock. A real-time algorithm for moving horizon state and parameter estimation. *Computers & chemical engineering*, 35(1):71–83, 2011.

- [66] Bokyu Kwon and Soohye Han. Least-mean-square receding horizon estimation. *Mathematical Problems in Engineering*, 2012, 2012.
- [67] Wook Hyun Kwon, Pyung Soo Kim, and PooGyeon Park. A receding horizon Kalman FIR filter for discrete time-invariant systems. *Automatic Control, IEEE Transactions on*, 44(9):1787–1791, 1999.
- [68] Wook Hyun Kwon, Pyung Soo Kim, and Soo Hee Han. A receding horizon unbiased fir filter for discrete-time state space models. *Automatica*, 38(3):545–551, 2002.
- [69] Frank L Lewis, Draguna Vrabie, and Vassilis L Syrmos. *Optimal control*. Wiley, 2012.
- [70] D Limon, I Alvarado, T Alamo, and EF Camacho. Robust tube-based MPC for tracking of constrained linear systems with additive disturbances. *Journal of Process Control*, 20(3):248–260, 2010.
- [71] Daniel Limón, Ignacio Alvarado, T Alamo, and Eduardo F Camacho. MPC for tracking piecewise constant references for constrained linear systems. *Automatica*, 44(9):2382–2387, 2008.
- [72] KV Ling and KW Lim. Receding horizon recursive state estimation. *Automatic Control, IEEE Transactions on*, 44(9):1750–1753, 1999.
- [73] R. López-Negrete, S.C. Patwardhan, and L.T. Biegler. Constrained particle filter approach to approximate the arrival cost in moving horizon estimation. *Journal of Process Control*, 21(6):909–919, 2011.
- [74] Jingran Ma, Joe Qin, Timothy Salsbury, and Peng Xu. Demand reduction in building energy systems based on economic model predictive control. *Chemical Engineering Science*, 67(1):92–100, 2012.

-
- [75] J. M. Maciejowski. *Predictive control with constraints*. Prentice-Hall, 2001.
- [76] J.B. Mare and J.A. de Dona. Symmetry between constrained reference tracking and constrained state estimation. *Automatica*, 45(1):207–211, 2009.
- [77] Jacob Mattingley and Stephen Boyd. Real-time convex optimization in signal processing. *Signal Processing Magazine, IEEE*, 27(3):50–61, 2010.
- [78] D.Q. Mayne, J.B. Rawlings, C.V. Rao, and P.O.M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000.
- [79] Richard J Meinhold and Nozer D Singpurwalla. Robustification of Kalman filter models. *Journal of the American Statistical Association*, 84(406):479–486, 1989.
- [80] Muhmmad J Mirza. A modified Kalman filter for non-Gaussian measurement noise. In *Communication Systems and Information Technology*, pages 401–409. Springer, 2011.
- [81] R V Mises. On the asymptotic distribution of differentiable statistical functions. *The annals of mathematical statistics*, 18(3):309–348, 1947.
- [82] Sorin Olaru and Didier Dumur. Compact explicit MPC with guarantee of feasibility for tracking. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 969–974, Seville, Spain, 2005.
- [83] G. Pannocchia, J.B. Rawlings, D.Q. Mayne, and W. Marquardt. On computing solutions to the continuous time constrained linear quadratic regulator . *Automatic Control, IEEE Transactions on*, 55(9):2192–2198, 2010.
- [84] Gabriele Pannocchia and Eric C Kerrigan. Offset-free receding horizon control of constrained linear systems. *AIChE Journal*, 51(12):3134–3146, 2005.

- [85] B. Pattipati, C. Sankavaram, K. Pattipati, Y. Zhang, M. Howell, and M. Salman. Multiple model moving horizon estimation approach to prognostics in coupled systems. In *AUTOTESTCON, 2011 IEEE*, pages 149–157, 2011.
- [86] T. Poloni, A.A. Eielsen, B. Rohal’-Ilkiv, and T.A. Johansen. Moving horizon observer for vibration dynamics with plant uncertainties in nanopositioning system estimation. In *American Control Conference (ACC), 2012*, pages 3817–3824, 2012.
- [87] S.J. Qin and T.A. Badgwell. A survey of industrial model predictive control technology. *Control engineering practice*, 11(7):733–764, 2003.
- [88] Christopher V Rao, James B Rawlings, and Jay H Lee. Constrained linear state estimationa moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.
- [89] C.V. Rao, J.B. Rawlings, and D.Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *Automatic Control, IEEE Transactions on*, 48(2):246–258, 2003.
- [90] K.J. Åström’s book and B. Wittenmark. *Computer-controlled systems: theory and design*, volume 3. prentice hall Englewood Cliffs, NJ, 1990.
- [91] James B Rawlings and Rishi Amrit. Optimizing process economic performance using model predictive control. In *Nonlinear model predictive control*, pages 119–138. Springer, 2009.
- [92] James B. Rawlings and David Q. Mayne. *Model Predictive Control: Theory and Design*. Nob Hill Publishing, 2009.

-
- [93] James B Rawlings, David Angeli, and Cuyler N Bates. Fundamentals of economic model predictive control. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 3851–3861, 2012.
- [94] Douglas G Robertson and Jay H Lee. On the use of constraints in least squares estimation and control. *Automatica*, 38(7):1113–1123, 2002.
- [95] Michael Roth, Emre Ozkan, and Fredrik Gustafsson. A student’s t filter for heavy tailed process and measurement noise. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5770–5774, 2013.
- [96] Francesco Scibilia, Sorin Olaru, and Morten Hovd. On feasible sets for MPC and their approximations. *Automatica*, 47(1):133–139, 2011.
- [97] Pierre OM Scokaert and James B Rawlings. Constrained linear quadratic regulation. *Automatic Control, IEEE Transactions on*, 43(8):1163–1169, 1998.
- [98] Pierre OM Scokaert, David Q Mayne, and James B Rawlings. Suboptimal model predictive control (feasibility implies stability). *Automatic Control, IEEE Transactions on*, 44(3):648–654, 1999.
- [99] Kurt W Simon and Allen R Stubberud. Duality of linear estimation and control. *Journal of Optimization Theory and Applications*, 6(1):55–67, 1970.
- [100] Eduardo D Sontag. *Mathematical control theory: deterministic finite dimensional systems*. Springer, 1998.
- [101] Florin Stoican, Sorin Olaru, María M Seron, and José A De Doná. Reference governor design for tracking problems with fault detection guarantees. *Journal of Process Control*, 22(5):829–836, 2012.

- [102] Jing Sun and Ilya V. Kolmanovsky. Load governor for fuel cell oxygen starvation protection: A robust nonlinear reference governor approach. *IEEE Transactions on Control Systems Technology*, 13(6):911–920, 2005.
- [103] Mario Sznaier and Mark J Damborg. Heuristically enhanced feedback control of constrained systems: the minimum time case. In *Decision and Control, 1991., Proceedings of the 30th IEEE Conference on*, pages 1730–1735, 1991.
- [104] D Wang and JA Romagnoli. A framework for robust data reconciliation based on a generalized objective function. *Industrial & engineering chemistry research*, 42(13):3075–3084, 2003.
- [105] D Wang and JA Romagnoli. Generalized t distribution and its applications to process data reconciliation and process monitoring. *Transactions of the Institute of Measurement and Control*, 27(5):367–390, 2005.
- [106] Liuping Wang. *Model Predictive Control System Design and Implementation Using MATLAB*. Springer, 2009.
- [107] Yang Wang, Rong Yang, Keck Voon Ling, and Eng Kee Poh. Robust vector tracking loop using moving horizon estimation. In *Proceedings of the ION 2015 Pacific PNT Meeting*, pages 640–648, 2015.
- [108] Roy E. Welsch. Influence functions and regression diagnostics. In R.L. Launer and A.F. Siegel, editors, *Modern data analysis*. Academic Press, New York, 1982.
- [109] Juan Yuz, Graham Goodwin, Arie Feuer, and Jose De Dona. Control of constrained linear systems using fast sampling rates. *Systems & control letters*, 54(10):981–990, 2005.

- [110] V.M. Zavala, C.D. Laird, and L.T. Biegler. A fast moving horizon estimation algorithm based on nonlinear programming sensitivity. *Journal of Process Control*, 18(9):876–884, 2008.
- [111] Günter M Ziegler. *Lectures on polytopes*. Springer, 1995.

Appendix

1 Proof of Theorem 2.1

Suppose $x_0 \in \mathbb{X}_0(N, \mathbb{X}_f)$ and the optimal control sequence is $\{u_k^*\}_{k=0}^{N-1}$. Apply the first input of the optimal sequence, u_0^* to the plant then the state is steered to $x^+ = Ax_0 + Bu_0^*$. Because the $\mathbb{X}_f$ is a invariant set under the controller $u = K_fx$ there exists an admissible control sequence $\left\{\{u_k^*\}_{k=1}^{N-1}, K_fx_N\right\}$ which steers the state from x^+ into $\mathbb{X}_f$ in finite number N of steps, while keeping the evolution of the state inside $\mathbb{X}$. So $x^+ \in \mathbb{X}_0(N, \mathbb{X}_f)$. Then the theorem is proved.

2 From Linear Quadratic Tracking to Kalman Filter by symmetry

Consider the following linear system

$$\begin{aligned} x_{k+1} &= A_c x_k + B_c u_k, \\ y_k &= C_c x_k, \\ e_k &= y_k^r - C_c x_k, \end{aligned} \tag{1}$$

where y_k^r is the reference signal.

Given x_i the LQT is defined as

$$\begin{aligned} P_c : \quad \min_{\{u_k\}_{k=i}^{N-1}} J_c^i &= \frac{1}{2} \sum_{k=i}^{N-1} \|u_k\|_{R_c}^2 + \frac{1}{2} \sum_{k=i}^{N-1} \|y_k^r - C_c x_k\|_{Q_c}^2 + \frac{1}{2} \|x_N - \mu_{cN}\|_{P_{cN}}^2, \\ \text{s.t. } x_{k+1} &= A_c x_k + B_c u_k, \end{aligned} \tag{2}$$

where $P_{cN} \geq 0$, $Q_c \geq 0$, $R_c > 0$ are all symmetric matrices and the μ_{cN} can be chosen as any vector satisfying $y_N^r = C_c \mu_{cN}$.

According to [69] the optimal solution of the LQT (2) is

$$K_k = \left(B_c' S_{k+1} B_c + R_c \right)^{-1} B_c' S_{k+1} A_c, \quad S_N = P_{cN}, \tag{3a}$$

$$S_k = A_c' S_{k+1} (A_c - B_c K_k) + C_c' Q_c C_c, \tag{3b}$$

$$l_k = (A_c - B_c K_k)' l_{k+1} + C_c' Q_c y_k^r, \quad l_N = P_{cN} \mu_{cN}, \tag{3c}$$

$$K_k^l = \left(B_c' S_{k+1} B_c + R_c \right)^{-1} B_c', \tag{3d}$$

$$u_k = -K_k x_k + K_k^l l_{k+1} \tag{3e}$$

where $k = i, i + 1, \dots, N - 1$, and the optimal value of J_c^i is:

$$J_c^{i*} = \frac{1}{2} x_i' S_i x_i - x_i' l_i + z_i, \quad (4)$$

where z_i does not depend on x_i .

Next consider the following observable system

$$\begin{aligned} x_{k+1} &= A_e x_k + B_e \omega_k, \\ y_k &= C_e x_k, \\ v_k &= y_k^d - C_e x_k, \end{aligned}$$

where y_k^d is the measurement. The process noise ω_k and measurement noise v_k are assumed zero mean and white noises with variances Q_e and R_e respectively. The initial state, with estimate μ_{ei} , an approximation mean, and the associated covariance matrix P_{ei} which is positive definite, is assumed to be uncorrelated with the two noise sequences.

The least-squares estimation is formulated as

$$\begin{aligned} P_e : \quad \min_{\hat{x}_i, \{\hat{\omega}_k\}_{k=i}^{N-1}} J_e^N &= \frac{1}{2} \|\hat{x}_i - \mu_{ei}\|_{P_{ei}^{-1}}^2 + \frac{1}{2} \sum_{k=i}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2, \\ s.t. \quad \hat{x}_{k+1} &= A_e \hat{x}_k + B_e \hat{\omega}_k, \end{aligned} \quad (5)$$

where $P_{ei} \geq 0$, $Q_e > 0$, $R_e \geq 0$ are all symmetric matrices.

The primal least-squares estimation (5) can be reformulated as

$$P_e : \quad \min_{\hat{x}_N} J_e^{N*}(\hat{x}_N) \quad (6)$$

$$J_e^{N*}(\hat{x}_N) = \min_{\{\hat{\omega}_k\}_{k=i}^{N-1}} J_e^N(\hat{x}_N, \{\hat{\omega}_k\}_{k=i}^{N-1}) \quad (7)$$

$$s.t. \quad \hat{x}_k = A_e^{-1} \hat{x}_{k+1} - A_e^{-1} B_e \hat{\omega}_k,$$

where

$$J_e^N(\hat{x}_N, \{\hat{\omega}_k\}_{k=i}^{N-1}) = \frac{1}{2} \|\hat{x}_i - \mu_{ei}\|_{P_{ei}^{-1}}^2 + \frac{1}{2} \sum_{k=i}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2.$$

The problem (7) can be reformulated in the reversing time form of the LQT (2) by variables and parameters translation in Table 1 and Table 2.

Table 1: Variables translation between LQT and Kalman filter

P_c		P_e	P_c		P_e	P_c		P_e
u_i	$\iff$	$\hat{\omega}_{N-1}$	x_i	$\iff$	$\hat{x}_N$	y_i^r	$\iff$	y_N^d
u_{i+1}	$\iff$	$\hat{\omega}_{N-2}$	x_{i+1}	$\iff$	$\hat{x}_{N-2}$	y_{i+2}^r	$\iff$	y_{N-2}^d
$\vdots$	$\iff$	$\vdots$	$\vdots$	$\iff$	$\vdots$	$\vdots$	$\iff$	$\vdots$
u_{N-1}	$\iff$	$\hat{\omega}_i$	x_N	$\iff$	$\hat{x}_i$	y_{N-1}^r	$\iff$	y_{i+1}^d

Table 2: Parameters translation between LQT and Kalman filter

P_c		P_e	P_c		P_e
A_c	$\iff$	A_e^{-1}	Q_c	$\iff$	R_e^{-1}
B_c	$\iff$	$-A_e^{-1}B_e$	R_c	$\iff$	Q_e^{-1}
C_c	$\iff$	C_e	μ_{cN}	$\iff$	μ_{ei}
P_{cN}	$\iff$	P_{ei}^{-1}			

Thus according to the solution of LQT, i.e., (3) and the symmetry relationship between LQT and least-squares estimation in Tables 1 and 2 we have

$$J_e^{N*}(\hat{x}_N) = \frac{1}{2} \hat{x}_N' S_N \hat{x}_N - \hat{x}_N' l_N + z_N$$

$$= \frac{1}{2} \|\hat{x}_N - S_N^{-1} l_N\|_{S_N}^2 - \frac{1}{2} l_N' S_N^{-1} l_N + z_N, \quad (8)$$

where

$$S_k = (A_e^{-1})' S_{k-1} (A_e^{-1} - (-A_e^{-1} B_e) K_k) + C_e' R_e^{-1} C_e, \quad S_i = P_{ei}^{-1}, \quad (9a)$$

$$K_k = \left((-A_e^{-1} B_e)' S_{k-1} (-A_e^{-1} B_e) + Q_e^{-1} \right)^{-1} (-A_e^{-1} B_e)' S_{k-1} A_e^{-1}, \quad (9b)$$

$$l_k = (A_e^{-1} - (-A_e^{-1} B_e) K_k)' l_{k-1} + C_e' R_e^{-1} y_k^d, \quad l_i = P_{ei}^{-1} \mu_{ei}. \quad (9c)$$

So the optimal solution of problem (6) is

$$\hat{x}_N = S_N^{-1} l_N. \quad (10)$$

The solution (9a) to (10) can be obtained from the estimation (5) using forward dynamic programming. Firstly we rearrange the overall objective function (7):

$$\begin{aligned} J_e^{N*}(\hat{x}_N) &= \min_{\{\hat{\omega}_k\}_{k=i}^{N-1}} \frac{1}{2} \|\hat{x}_i - \mu_{ei}\|_{P_{ei}^{-1}}^2 + \frac{1}{2} \sum_{k=i}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2 \\ &= \min_{\hat{\omega}_i} \frac{1}{2} \|\hat{x}_i - \mu_{ei}\|_{P_{ei}^{-1}}^2 + \|\hat{\omega}_i\|_{Q_e^{-1}}^2 \\ &\quad + \min_{\{\hat{\omega}_k\}_{k=i+1}^{N-1}} \frac{1}{2} \sum_{k=i+1}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2 \\ &= \min_{\hat{\omega}_i} \frac{1}{2} \|A_e^{-1} \hat{x}_{i+1} + (-A_e^{-1} B_e) \hat{\omega}_i - \mu_{ei}\|_{P_{ei}^{-1}}^2 + \|\hat{\omega}_i\|_{Q_e^{-1}}^2 \\ &\quad + \min_{\{\hat{\omega}_k\}_{k=i+1}^{N-1}} \frac{1}{2} \sum_{k=i+1}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2 \\ &= \min_{\hat{\omega}_i} \frac{1}{2} \|\hat{\omega}_i - \hat{\omega}_i^*\|_H^2 + F \\ &\quad + \min_{\{\hat{\omega}_k\}_{k=i+1}^{N-1}} \frac{1}{2} \sum_{k=i+1}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2, \end{aligned} \quad (11)$$

where

$$\begin{aligned}
H &= Q_e^{-1} + (-A_e^{-1}B_e)'P_{ei}^{-1}(-A_e^{-1}B_e) \\
\hat{\omega}_i^* &= H^{-1}(-A_e^{-1}B_e)'P_{ei}^{-1}(\mu_{ei} - A_e^{-1}\hat{x}_{i+1}), \\
&= -\left(Q_e^{-1} + (-A_e^{-1}B_e)'S_i(-A_e^{-1}B_e)\right)^{-1}(-A_e^{-1}B_e)'S_iA_e^{-1}\hat{x}_{i+1} \\
&\quad + \left(Q_e^{-1} + (-A_e^{-1}B_e)'S_i(-A_e^{-1}B_e)\right)^{-1}(-A_e^{-1}B_e)'S_i\mu_{ei} \\
&= -K_{i+1}\hat{x}_{i+1} + K_{i+1}^l l_i, \\
F &= \frac{1}{2}\|A_e^{-1}\hat{x}_{i+1} + (-A_e^{-1}B_e)\hat{\omega}_i^* - \mu_{ei}\|_{S_i}^2 + \|\hat{\omega}_i^*\|_{Q_e^{-1}}^2,
\end{aligned} \tag{12}$$

and

$$\begin{aligned}
S_i &= P_{ei}^{-1}, \\
K_{i+1} &= \left(Q_e^{-1} + (-A_e^{-1}B_e)'S_i(-A_e^{-1}B_e)\right)^{-1}(-A_e^{-1}B_e)'S_iA_e^{-1}, \\
K_{i+1}^l &= \left(Q_e^{-1} + (-A_e^{-1}B_e)'S_i(-A_e^{-1}B_e)\right)^{-1}(-A_e^{-1}B_e)', \\
l_i &= S_i\mu_{ei}.
\end{aligned} \tag{13}$$

Substituting $\hat{\omega}_i = \hat{\omega}_i^*$ into (11) we have

$$\begin{aligned}
J_e^{N*}(\hat{x}_N) &= \min_{\{\hat{\omega}_k\}_{k=i+1}^{N-1}} F + \frac{1}{2} \sum_{k=i+1}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^N \|y_k^d - C_e x_k\|_{R_e^{-1}}^2 \\
&= \min_{\{\hat{\omega}_k\}_{k=i+1}^{N-1}} \frac{1}{2} \|A_e^{-1}\hat{x}_{i+1} + (-A_e^{-1}B_e)(-K_{i+1}\hat{x}_{i+1} + K_{i+1}^l l_i) - S_i^{-1}l_i\|_{S_i}^2 \\
&\quad + \|-K_{i+1}\hat{x}_{i+1} + K_{i+1}^l l_i\|_{Q_e^{-1}}^2 \\
&\quad + \frac{1}{2} \|y_{i+1}^d - C_e \hat{x}_{i+1}\|_{R_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+1}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+2}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2 \\
&= \min_{\hat{\omega}_{i+1}} \frac{1}{2} \|\hat{x}_{i+1} - g_{i+1}\|_{S_{i+1}}^2 + \frac{1}{2} \|\hat{\omega}_{i+1}\|_{Q_e^{-1}}^2 + \min_{\{\hat{\omega}_k\}_{k=i+2}^{N-1}} \frac{1}{2} \sum_{k=i+2}^N \|y_k^d - C_e \hat{x}_k\|_{R_e^{-1}}^2
\end{aligned}$$

$$+ \frac{1}{2} \sum_{k=i+2}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 + d_{i+1}, \quad (14)$$

where

$$\begin{aligned} S_{i+1} &= (A_e^{-1} - (-A_e^{-1}B_e)K_{i+1})' S_i (A_e^{-1} - (-A_e^{-1}B_e)K_{i+1}) + K_{i+1}' Q_e^{-1} K_{i+1} + C_e' R_e^{-1} C_e \\ &= (A_e^{-1})' S_i (A_e^{-1} - (-A_e^{-1}B_e)K_{i+1}) + C_e' R_e^{-1} C_e, \\ g_{i+1} &= S_{i+1}^{-1} \left[(A_e^{-1} - (-A_e^{-1}B_e)K_{i+1})' S_i (S_i^{-1} l_i - (-A_e^{-1}B_e)K_{i+1}^l l_i) + K_{i+1}' Q_e^{-1} K_{i+1}^l l_N \right. \\ &\quad \left. + C_e' R_e^{-1} y_{i+1}^d \right]. \end{aligned}$$

Setting the $d_{i+1} = 0$ which does not depend on $\hat{x}_{i+1}$.

$$\begin{aligned} J_e^{N*}(\hat{x}_N) &= \min_{\hat{\omega}_{i+1}} \frac{1}{2} \|A_e^{-1} x_{i+2} + (-A_e^{-1}B_e)\hat{\omega}_{i+1} - g_{i+1}\|_{S_{i+1}} + \frac{1}{2} \|\hat{\omega}_{i+1}\|_{Q_e^{-1}}^2 \\ &\quad + \min_{\{\hat{\omega}_k\}_{k=i+2}^{N-1}} \frac{1}{2} \sum_{k=i+2}^N \|y_k^d - C_e x_k\|_{R_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+2}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2 \\ &= \min_{\hat{\omega}_{i+1}} \frac{1}{2} \|\hat{\omega}_{i+1} - \hat{\omega}_{i+1}^*\|_H \\ &\quad + \min_{\{\hat{\omega}_k\}_{k=i+2}^{N-1}} F + \frac{1}{2} \sum_{k=i+2}^N \|y_k^d - C_e x_k\|_{R_e^{-1}}^2 + \frac{1}{2} \sum_{k=i+2}^{N-1} \|\hat{\omega}_k\|_{Q_e^{-1}}^2, \quad (15) \end{aligned}$$

where

$$\begin{aligned} H &= Q_e^{-1} + (-A_e^{-1}B_e)' S_{i+1} (-A_e^{-1}B_e), \\ \hat{\omega}_{i+1}^* &= H^{-1} (-A_e^{-1}B_e)' S_{i+1} (g_{i+1} - A_e^{-1} \hat{x}_{i+2}) \\ &= - (Q_e^{-1} + (-A_e^{-1}B_e)' S_{i+1} (-A_e^{-1}B_e))^{-1} (-A_e^{-1}B_e)' S_{i+1} A_e^{-1} x_{N-2} \\ &\quad + (Q_e^{-1} + (-A_e^{-1}B_e)' S_{i+1} (-A_e^{-1}B_e))^{-1} (-A_e^{-1}B_e)' S_{i+1} g_{i+1} \\ &= - K_{i+2} \hat{x}_{i+2} + K_{i+2}^l l_{i+1}, \\ F &= \frac{1}{2} \|A_e^{-1} \hat{x}_{i+2} + (-A_e^{-1}B_e)\hat{\omega}_{i+1}^* - g_{i+1}\|_{S_{i+1}} + \frac{1}{2} \|\hat{\omega}_{i+1}^*\|_{Q_e^{-1}}^2, \\ K_{i+2} &= (Q_e^{-1} + (-A_e^{-1}B_e)' S_{i+1} (-A_e^{-1}B_e))^{-1} (-A_e^{-1}B_e)' S_{i+1} A_e^{-1}, \end{aligned}$$

$$\begin{aligned}
K_{i+2}^l &= (Q_e^{-1} + (-A_e^{-1}B_e)'S_{i+1}(-A_e^{-1}B_e))^{-1}(-A_e^{-1}B_e)', \\
l_{i+1} &= (A_e^{-1} - (-A_e^{-1}B_e)K_{i+1})'S_i(S_i^{-1}l_i - (-A_e^{-1}B_e)K_{i+1}^l l_i) \\
&\quad + K_{i+1}'Q_e^{-1}K_{i+1}^l l_i + C_e'R_e^{-1}y_{i+1}^d \\
&= (A_e^{-1} - (-A_e^{-1}B_e)K_{i+1})'l_i + C_e'R_e^{-1}y_{i+1}^d.
\end{aligned} \tag{16}$$

Proceed to use the forward dynamic programming until we get $\hat{\omega}_{N-1}^*$ and $J_e^{N*}(\hat{x}_N)$:

$$\begin{aligned}
\hat{\omega}_{N-1}^* &= -K_N \hat{x}_N + K_N^l l_{N-1}, \\
J_e^{N*}(\hat{x}_N) &= \frac{1}{2} \|\hat{x}_N - S_N^{-1} l_N\|_{S_N} + d_N,
\end{aligned} \tag{17}$$

where K_{N-1} , K_{N-1}^l , l_N and S_N can be obtained by the following iterative equations:

$$\begin{aligned}
S_{k+1} &= (A_e^{-1})' S_k (A_e^{-1} - (-A_e^{-1}B_e)K_{k+1}) + C_e'R_e^{-1}C_e \quad S_i = P_{ei}^{-1}, \\
K_{k+1} &= (Q_e^{-1} + (-A_e^{-1}B_e)'S_k(-A_e^{-1}B_e))^{-1}(-A_e^{-1}B_e)'S_k A_e^{-1}, \\
K_{k+1}^l &= (Q_e^{-1} + (-A_e^{-1}B_e)'S_k(-A_e^{-1}B_e))^{-1}(-A_e^{-1}B_e)', \\
l_{k+1} &= (A_e^{-1} - (-A_e^{-1}B_e)K_{k+1})'l_k + C_e'R_e^{-1}y_{k+1}^d, \quad l_i = S_i \mu_{ei}
\end{aligned} \tag{18}$$

which are (9). In (17) d_N does not depend on $\hat{x}_N$. So the optimal solution of (6) is

$$\hat{x}_N = S_N^{-1} l_N.$$

3 Direct Method to compute the set of admissible measurements of MHE (4.13a)

Based on the model (4.12) we obtain

$$\begin{aligned} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix} &= \begin{bmatrix} C_e A_e \\ C_e A_e^2 \\ \vdots \\ C_e A_e^N \end{bmatrix} \hat{x}_0 + \begin{bmatrix} C_e B_e & 0 & 0 & \cdots & 0 \\ C_e A_e B_e & C_e B_e & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_e A_e^{N-1} B_e & C_e A_e^{N-2} B_e & C_e A_e^{N-3} B_e & \cdots & C_e B_e \end{bmatrix} \begin{bmatrix} \hat{\omega}_0 \\ \hat{\omega}_1 \\ \vdots \\ \hat{\omega}_{N-1} \end{bmatrix} \\ &= E \hat{x}_0 + Z W_N, \end{aligned}$$

where $W_N = \begin{bmatrix} \hat{\omega}'_0 & \hat{\omega}'_1 & \cdots & \hat{\omega}'_{N-1} \end{bmatrix}'$. So

$$\Omega_{\hat{y}_k} = \text{range}(\Omega_{x_0}, E, \mathbf{0}) \oplus \text{range}(\Omega_{W_N}, Z, \mathbf{0}), \quad (19)$$

where $\Omega_{W_N} = [\Omega_{\omega}, \dots, \Omega_{\omega}]'$.

With

$$\begin{bmatrix} y_1^d \\ y_2^d \\ \vdots \\ y_N^d \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix} + \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \vdots \\ \hat{v}_N \end{bmatrix} \quad (20)$$

we have

$$\Omega_{y_k^d} = \Omega_{\hat{y}_k} \oplus \Omega_{V_N}, \quad (21)$$

where $\Omega_{V_N} = [\Omega_v, \dots, \Omega_v]'$. Substituting equation (19) to (21)

$$\Omega_{y_k^d} = \text{range}(\Omega_{x_0}, E, \mathbf{0}) \oplus \text{range}(\Omega_{W_N}, Z, \mathbf{0}) \oplus \Omega_{V_N}. \quad (22)$$

4 Derivation of (5.10)

Differentiating Equation (5.9) wrt. h gives

$$\sum_{k=1}^T \int_{-\infty}^{\infty} \frac{d\psi_k(e_k)}{dh} [(1-h)f(e_k) + h\Delta_{e_k}] de_k + \sum_{k=1}^T \int_{-\infty}^{\infty} \psi_k(e_k) [-f(e_k) + \Delta_{e_k}] de_k = 0. \quad (23)$$

Substituting

$$\frac{d\psi_k(e_k)}{dh} = \frac{d\psi_k(e_k)}{d\kappa(h)} \frac{d\kappa(h)}{dh} \quad \text{and} \quad \int_{-\infty}^{\infty} \psi_k(e_k) f(e_k) de_k = 0$$

to (23) gives

$$\sum_{k=1}^T \int_{-\infty}^{\infty} \frac{d\psi_k(e_k)}{d\kappa(h)} [(1-h)f(e_k) + h\Delta_{e_k}] de_k \frac{d\kappa(h)}{dh} + \sum_{k=1}^T \int_{-\infty}^{\infty} \psi_k(e_k) \Delta_{e_k} de_k = 0. \quad (24)$$

Setting $h = 0$ in (24) gives

$$\sum_{k=1}^T \int_{-\infty}^{\infty} \frac{d\psi_k(e_k)}{d\kappa(h)} f(e_k) de_k \frac{d\kappa(h)}{dh} \Big|_{h=0} + \sum_{k=1}^T \int_{-\infty}^{\infty} \psi_k(e_k) \Delta_{e_k} de_k = 0.$$

Recall that in (5.5) $e_k = y_k - (H\Phi^{k-1})' x_1 - H\bar{x}_k$ thus e_k at $h = 0$ is equal to $y_k - (H\Phi^{k-1})' \kappa(0) - H\bar{x}_k$. We denote $\kappa(0)$ as $\bar{x}_1$.

Thus we have

$$\frac{d\kappa(h)}{dh} \Big|_{h=0} = - \left(\sum_{k=1}^T \int_{-\infty}^{\infty} \frac{d\psi_k(e_k)}{dx_1} f(e_k) de_k \right)^{-1} \sum_{k=1}^T \psi_k(e_k) \Big|_{x_1=\bar{x}_1}. \quad (25)$$

Changing the integration variable e_k with e in (25) gives

$$\frac{d\kappa(h)}{dh} \Big|_{h=0} = - \left(\sum_{k=1}^T \int_{-\infty}^{\infty} \frac{d\psi_k(e)}{dx_1} f(e) de \right)^{-1} \sum_{k=1}^T \psi_k(e_k) \Big|_{x_1=\bar{x}_1}.$$

which is (5.10).

5 Proof of Proposition 5.1

Substituting the t -distribution noise pdf (5.1) to J

$$\begin{aligned}\ln f(e_k) &= -\ln \left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\sigma\Gamma(\frac{\nu}{2})} \left(1 + \frac{(y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)^2}{\sigma^2\nu} \right)^{-\frac{\nu+1}{2}} \right) \\ &= -\ln \left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\sigma\Gamma(\frac{\nu}{2})} \right) + \frac{\nu+1}{2} \ln \left(1 + \frac{(y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)^2}{\sigma^2\nu} \right).\end{aligned}$$

where J is the cost function of MLE for the ARMAX process with t -distribution noise. It is the sum of the log of the likelihood function, i.e., $f(e)$, given the measurements up to current time instant T .

Differentiating $\ln f(e_k)$ with respect to x_1 gives

$$\begin{aligned}\psi_k(e_k) &= \frac{d}{dx} (\ln f(e_k)) = \frac{\nu+1}{2} \frac{d}{dx_1} \left(\ln \left(1 + \frac{(y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)^2}{\sigma^2\nu} \right) \right) \\ &= -\frac{(\nu+1)(H\Phi^{k-1})' (y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)}{\sigma^2\nu + (y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)^2} = -\frac{(\nu+1)(H\Phi^{k-1})' e_k}{\sigma^2\nu + (e_k)^2}.\end{aligned}\tag{26}$$

Differentiating $\psi_k(e)$ with respect to x_1

$$\frac{d\psi_k(e)}{dx_1} = -\frac{d}{dx_1} \left(\frac{(\nu+1)(H\Phi^{k-1})' e}{\sigma^2\nu + e^2} \right) = (H\Phi^{k-1})' H\Phi^{k-1} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2}.\tag{27}$$

Substituting (26) and (27) to (5.10) gives

$$\text{IF}(e) = \left(\sum_{k=1}^T \int_{-\infty}^{\infty} (H\Phi^{k-1})' (H\Phi^{k-1}) \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1}$$

$$\times \sum_{k=1}^T \frac{(\nu+1) (H\Phi^{k-1})' e_k}{\sigma^2\nu + (e_k)^2} \Big|_{x_1=\bar{x}_1}$$

which is (5.13).

6 Proof of Corollary 5.1

When $\nu = \infty$ by using L'Hopital's rule twice we have

$$\begin{aligned}
 & \lim_{\nu \rightarrow \infty} \int_{-\infty}^{\infty} \frac{(\nu + 1) (H\Phi^{k-1})' H\Phi^{k-1} (\sigma^2 \nu - e^2)}{(\sigma^2 \nu + e^2)^2} f(e) de \\
 &= (H\Phi^{k-1})' H\Phi^{k-1} \lim_{\nu \rightarrow \infty} \int_{-\infty}^{\infty} \frac{2\sigma^2 \nu - e^2 + \sigma^2}{2\sigma^2 (\sigma^2 \nu + e^2)} f(e) de \\
 &= \frac{(H\Phi^{k-1})' H\Phi^{k-1}}{\sigma^2}.
 \end{aligned} \tag{28a}$$

In addition,

$$\begin{aligned}
 & \lim_{\nu \rightarrow \infty} \frac{(\nu + 1) (H\Phi^{k-1})' (y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)}{\sigma^2 \nu + (y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)^2} \\
 &= \frac{(H\Phi^{k-1})' (y_k - H\Phi^{k-1}x_1 - H\bar{x}_k)}{\sigma^2}, \quad k = 1, 2, \dots, T.
 \end{aligned} \tag{28b}$$

Substituting (28) to (5.13) gives

$$\text{IF}(e) = -\bar{x}_1 + \left(\sum_{k=1}^T (H\Phi^{k-1})' H\Phi^{k-1} \right)^{-1} \sum_{k=1}^T (H\Phi^{k-1})' (y_k - H\bar{x}_k)$$

which is (5.14). Then (5.15) is obtained by substituting (5.14) to (5.12).

7 Proof of Proposition 5.2

Rewriting the state space model (5.2) of ARMAX as

$$\begin{aligned} x_{k+1} &= \Phi x_k + \Gamma u_k + \Omega y_k \\ y_k &= H x_k + e_k \end{aligned} \quad (29)$$

where $\Phi = \Phi_a - \Omega H$. Then with $\hat{x}_1 = \text{IF}(e)$ the current state estimate of x_T is obtained by recursively applying the model (29)

$$\hat{x}_T = \Phi^{T-1} x_{\text{IF}} + \sum_{k=1}^{T-1} \Phi^{T-k-1} (\Gamma u_k + \Omega y_k). \quad (30)$$

Since

$$x_{\text{IF}} = \text{IF}(e) \quad (31)$$

(30) becomes

$$\begin{aligned} \hat{x}_T &= \Phi^{T-1} \text{IF}(e) + \sum_{k=1}^{T-1} \Phi^{T-k-1} (\Gamma u_k + \Omega y_k) \\ &= \bar{x}_T + \Phi^{T-1} \text{IF}(e), \end{aligned} \quad (32)$$

where

$$\bar{x}_T = \sum_{k=1}^{T-1} \Phi^{T-k-1} (\Gamma u_k + \Omega y_k).$$

With (5.18)

$$\Phi^{T-1} \text{IF}(e) = \left(\sum_{i=1}^N (H \Phi^{i-N})' H \Phi^{i-N} \right)^{-1} \Psi \mathcal{W}_{T,N}, \quad (33)$$

where

$$\begin{aligned}
 \Psi &= \begin{bmatrix} (H\Phi^{1-N})' & (H\Phi^{2-N})' & \cdots & H' \end{bmatrix}, \\
 \mathcal{W}_{T,N} &= \begin{bmatrix} w_{T-N+1} & w_{T-N+2} & \cdots & w_T \end{bmatrix}', \\
 w_{T-N+i} &= \left(\int_{-\infty}^{\infty} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \frac{(\nu+1)(y_{T-N+i} - H\bar{x}_{T-N+i})}{\sigma^2\nu + (y_{T-N+i} - H\bar{x}_{T-N+i})^2}, \\
 &\quad i = 1, 2, \dots, N.
 \end{aligned}$$

8 Proof of Theorem 5.1

Define

$$\Delta x_T = \left(\sum_{i=1}^N (H\Phi^{i-N})' H\Phi^{i-N} \right)^{-1} \Psi \mathcal{W}_{T,N}. \quad (34)$$

Thus the estimator (5.19a) can be written as

$$\hat{x}_T = \bar{x}_T + \Delta x_T.$$

Notice that Equation (34) gives Δx_T from the minimization of the Moving Window Least-Squares loss function

$$V = \frac{1}{2} \sum_{i=1}^N (w_{T-N+i} - H\Phi^{i-N} \Delta x_T)^2.$$

The recursive solution of the above Moving Window Least-Squares [72] is

$$\Delta x_{T+1} = (\Phi - L_N H\Phi) \Delta x_T + L_N w_{T+1} - \tilde{L}_N (w_{T-N+1} - H\Phi^{-N+1} \Delta x_T),$$

where

$$\begin{aligned} L_N &= \Phi^{N-1} P_N (H\Phi^{N-1})', \quad \tilde{L}_N = \Phi^{N-1} P_N (H\Phi^{-1})', \\ P_N &= \left(\sum_{i=1}^N (H\Phi^{i-1})' H\Phi^{i-1} \right)^{-1}. \end{aligned}$$

In the next instant

$$\begin{aligned} \hat{x}_{T+1} &= \bar{x}_{T+1} + \Delta x_{T+1} = \Phi \bar{x}_T + \Gamma u_T + \Omega y_T + (\Phi - L_N H\Phi) \Delta x_T + L_N w_{T+1} \\ &\quad - \tilde{L}_N (w_{T-N+1} - H\Phi^{-N+1} \Delta x_T) \\ &= \Phi \hat{x}_T + \Gamma u_T + \Omega y_T + L_N (w_{T+1} - H\Phi \Delta x_T) \\ &\quad - \tilde{L}_N (w_{T-N+1} - H\Phi^{-N+1} \Delta x_T). \end{aligned} \quad (35)$$

Recall that in (5.3)

$$\bar{x}_T = \sum_{i=1}^{T-1} \Phi^{T-i-1} (\Gamma u_i + \Omega y_i)$$

which gives

$$\bar{x}_{T+1} = \Phi x_T + \Gamma u_T + \Omega y_T \quad (36)$$

and

$$\bar{x}_T = \Phi^{N-1} \bar{x}_{T-N+1} + \sum_{i=1}^{N-1} \Phi^{i-1} (\Gamma u_{T-i} + \Omega y_{T-i}).$$

So

$$\bar{x}_{T-N+1} = \Phi^{-N+1} \left(\bar{x}_T - \sum_{i=1}^{N-1} \Phi^{i-1} (\Gamma u_{T-i} + \Omega y_{T-i}) \right). \quad (37)$$

Define

$$z_k = w_k + H \bar{x}_k, \quad k = T - N + 1, T + 1. \quad (38)$$

Replacing w_{T+1} with $z_{T+1} - H \bar{x}_{T+1}$ and w_{T-N+1} with $z_{T-N+1} - H \bar{x}_{T-N+1}$ in (35) gives

$$\begin{aligned} \hat{x}_{T+1} &= \Phi \hat{x}_T + \Gamma u_T + \Omega y_T + L_N (z_{T+1} - H \bar{x}_{T+1} - H \Phi \Delta x_T) \\ &\quad - \tilde{L}_N (z_{T-N+1} - H \bar{x}_{T-N+1} - H \Phi^{-N+1} \Delta x_T). \end{aligned} \quad (39)$$

Substituting (36) and (37) to (39) then using the facts of $\hat{x}_{T+1} = \bar{x}_{T+1} + \Delta x_{T+1}$ and $\hat{x}_{T-N+1} = \bar{x}_{T-N+1} + \Delta x_{T-N+1}$ gives

$$\begin{aligned} \hat{x}_{T+1} &= \Phi \hat{x}_T + \Gamma u_T + \Omega y_T + L_N (z_{T+1} - H (\Phi \hat{x}_T + \Gamma u_T + \Omega y_T)) \\ &\quad - \tilde{L}_N \left(z_{T-N+1} - H \Phi^{-N+1} \hat{x}_T + H \Phi^{-N+1} \sum_{i=1}^{N-1} \Phi^{i-1} (\Gamma u_{T-i} + \Omega y_{T-i}) \right), \end{aligned} \quad (40)$$

Rewrite (40) as

$$\begin{aligned}\hat{x}_{T+1} = & \Phi\hat{x}_T + \Gamma u_T + \Omega y_T + L_N (z_{T+1} - H(\Phi\hat{x}_T + \Gamma u_T + \Omega y_T)) \\ & - \tilde{L}_N (z_{T-N+1} - H\Phi^{-N+1}\hat{x}_T + H\Phi^{-N+1}(\Xi_u U_{T-1,N-1} + \Xi_y Y_{T-1,N-1}))\end{aligned}$$

which is Equation (5.24a).

9 Proof of Theorem 5.2

From the estimator (5.19a) we have

$$E(\hat{x}_T) = E(\bar{x}_T) + \int_{-\infty}^{\infty} \Phi^{T-1} \text{IF}(e) g(e) de. \quad (41)$$

According to [47] we have

$$\int_{-\infty}^{\infty} \text{IF}(e) g(e) de = 0. \quad (42)$$

Rewrite $\bar{x}_T$ as

$$\bar{x}_T = \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} (\Gamma u_k + \Omega e_k). \quad (43)$$

Because e_k is zero mean

$$E(\bar{x}_T) = \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k. \quad (44)$$

Substituting (42) and (44) to (41) gives

$$E(\hat{x}_T) = \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} (\Gamma u_k + \Omega e_k) \quad (45)$$

which is equation (5.27a).

With (43) rewrite the estimator (5.19a) as

$$\hat{x}_T = \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} (\Gamma u_k + \Omega e_k) + \Phi^{T-1} \text{IF}(e). \quad (46)$$

Thus

$$\hat{x}_T - E(\hat{x}_T) = \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Omega e_k + \Phi^{T-1} \text{IF}(e).$$

As the pdf of e_k is $g(e)$, the variance of $\hat{x}_T$ is given by

$$\begin{aligned} Var(\hat{x}_T) &= \int_{-\infty}^{\infty} \left(\sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Omega e_k + \Phi^{T-1} \text{IF}(e) \right) \\ &\quad \times \left(\sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Omega e_k + \Phi^{T-1} \text{IF}(e) \right)' g(e) de. \end{aligned} \quad (47)$$

Substituting the $\text{IF}(e)$ of (5.18) to (47) where e_k is independent to each other and simplifying give

$$\begin{aligned} Var(\hat{x}_T) &= \frac{\rho_1}{\rho_4^2} M_N + 2 \sum_{i=1}^{N-1} \frac{\rho_2}{\rho_4} M_N (H \Phi^{i-N})' \left((\Phi + \Omega H)^{N-i-1} \Omega \right)' \\ &\quad + \sum_{k=1}^{T-1} \rho_3 \left((\Phi + \Omega H)^{T-k-1} \Omega \right) \left((\Phi + \Omega H)^{T-k-1} \Omega \right)', \end{aligned}$$

where

$$\begin{aligned} \rho_1 &= \int_{-\infty}^{\infty} \left(\frac{(\nu+1)e}{\sigma^2\nu + e^2} \right)^2 g(e) de, \quad \rho_2 = \int_{-\infty}^{\infty} \frac{(\nu+1)e^2}{\sigma^2\nu + e^2} g(e) de, \\ \rho_3 &= \int_{-\infty}^{\infty} e^2 g(e) de, \quad \rho_4 = \int_{-\infty}^{\infty} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} g(e) de, \\ M_N &= \left(\sum_{i=1}^N (H \Phi^{i-N})' H \Phi^{i-N} \right)^{-1}. \end{aligned}$$

10 Proof of Corollary 5.3

Recall that

$$\bar{x}_T = \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} (\Gamma u_k + \Omega e_k). \quad (48)$$

Because e_k is zero mean except that e_{k_1} at k_1 is not zero mean

$$E(\bar{x}_T) = \begin{cases} \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k & \text{if } T \leq k_1, \\ \sum_{k=1}^{T-1} (\Phi + \Omega H)^{T-k-1} \Gamma u_k + (\Phi + \Omega H)^{T-k_1-1} \Omega e_{k_1} & \text{if } T > k_1, \end{cases}$$

which is (5.34). Recall that

$$E(\hat{y}_T) = HE(\bar{x}_T) + H \int_{-\infty}^{\infty} \Phi^{T-1} \text{IF}(e) g(e) de. \quad (49)$$

In the second term of (49) the noise distribution $g(e)$ at k_1 is $\Delta_{e_{k_1}}$ because the outlier e_1 happens at k_1 . Thus

$$\begin{aligned} & H \int_{-\infty}^{\infty} \Phi^{T-1} \text{IF}(e) g(e) de \\ = & \begin{cases} HM_N \left(\int_{-\infty}^{\infty} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \int_{-\infty}^{\infty} \frac{(\nu+1)(H\Phi^{k_1-T})' e}{\sigma^2\nu + e^2} \Delta_{e_{k_1}} de \\ \quad - HM_N \left(\int_{-\infty}^{\infty} \frac{(\nu+1)(\sigma^2\nu - e^2)}{(\sigma^2\nu + e^2)^2} f(e) de \right)^{-1} \\ \quad \times \int_{-\infty}^{\infty} \sum_{\substack{i=1 \\ i \neq k_1 - T + N}}^N \frac{(\nu+1)(H\Phi^{i-1})' e}{\sigma^2\nu + e^2} g(e) de & \text{if } k_1 \leq T \leq k_1 + N - 1, \\ 0 & \text{if } T < k_1 \text{ or } T > k_1 + N - 1. \end{cases} \end{aligned} \quad (50)$$

With

$$-\int_{-\infty}^{\infty} \sum_{\substack{i=1 \\ i \neq k_1 - T + N}}^N \frac{(\nu + 1) (H\Phi^{i-1})' e}{\sigma^2 \nu + e^2} g(e) de = 0,$$

equation (50) becomes

$$\begin{aligned} & H \int_{-\infty}^{\infty} \Phi^{T-1} \text{IF}(e) g(e) de \\ &= \begin{cases} -HM_N \left(\int_{-\infty}^{\infty} \frac{(\nu + 1) (\sigma^2 \nu - e^2)}{(\sigma^2 \nu + e^2)^2} f(e) de \right)^{-1} \\ \quad \times \int_{-\infty}^{\infty} \frac{(\nu + 1) (H\Phi^{k_1-T})' e}{\sigma^2 \nu + e^2} \Delta_{e_{k_1}} de & \text{if } k_1 \leq T \leq k_1 + N - 1, \\ 0 & \text{if } T < k_1 \text{ or } T > k_1 + N - 1, \end{cases} \\ &= \begin{cases} \frac{1}{\rho_4} HM_N \frac{(H\Phi^{k_1-T})' (\nu + 1) e_{k_1}}{\sigma^2 \nu + e_{k_1}^2} & \text{if } k_1 \leq T \leq k_1 + N - 1 \\ 0 & \text{if } T < k_1 \text{ or } T > k_1 + N - 1. \end{cases} \end{aligned} \quad (51)$$

Equation (5.33) is obtained by substituting (51) to (49).