Moving Horizon State Estimation of Discrete Time Systems

by

Peter Klaus Findeisen

A thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Science

(Chemical Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON

1997

To my Parents

Acknowledgments

I would like to express my thanks and appreciation to my advisor, Dr. James B. Rawlings. I thoroughly enjoyed working together with him and would wish everybody to have such an insightful and inspiring advisor. It is due to his patience and guidance that I could reach my goal. I would further like to say thanks to the following people.

Professor M. Zeitz who made it possible so that I could take part on, in my opinion, a very successful exchange program between the Universität Stuttgart and the University of Wisconsin–Madison. I hope that this program will be also offered in the future, in order to allow other students to enjoy the great facilities and possibilities here in Madison.

I would also like to express my thanks to all students and former student of the Rawlings research group, who provided a great working and social environment: John Eaton, who created a excellent computing environment, who never complained about my sometimes bothering questions. John Campbell who shared the apartment with me during the final phase of my writing. Thanks John it was an great time. Raoul Bindlish who helped out in a time of great need and who encouraged me during the last few weeks of my stay in Madison; Rock Matthews with whom I had many interesting discussions from an wide field of topics; Christopher Rao who is an great person to talk to about an wide field of academic problems and who also offered a lot of help to me during the final phase of my writing; Daniel Patience the hard working men who was a lot of times the last person working (except the people writing there thesises); Hazim Mohameed who helped me out in a moment of great need at the day of my defense; Jenny Wang who "just" started enjoying to be an member of the JBR group before I left; Scott Middlebrooks and Jianfeng Zhang.

PETER KLAUS FINDEISEN

University of Wisconsin–Madison
July 1997

Moving Horizon State Estimation of Discrete Time Systems

Peter Klaus Findeisen
Under the supervision of Professor
James B. Rawlings
At the University of Wisconsin–Madison

This work considers the moving horizon estimation (MHE) problem. Most of todays applied or developed control schemes assume that the state of the controlled system is known explicitly. In reality the system states often cannot be measured directly; instead, they must be estimated from available output measurements. The traditional state estimation approach for linear systems under the influence of noise is the Kalman Filter (KF). However, this approach does not allow to consider constraints on the noise terms or the states, that could improve the estimates. Also the extension to nonlinear systems is not possible without linearization or other approximations.

A class of estimation methods that seem to offer solutions to these problems is the moving horizon estimation (MHE) methods. These methods can be motivated as the dual formulation of model predictive control (MPC) for state estimation. Most of the proposed MHE schemes place special assumptions on the weights and initial values in order to guarantee stability. Often this is done to allow the use of stability results of the KF. This work tries to establish stability conditions without using the corresponding KF results, thus allowing a broader and wider variety of weighting matrices and initial values.

This work offers contributions in three different areas. It contains an extensive review of the existing moving horizon estimation concepts for linear and nonlinear systems. It shows that most stability proofs for linear MHE schemes are based on KF properties. If nonlinear systems are considered this approach is in general not possible, which explains the existence of so few NMHE algorithms with stability guarantee.

The second part derives and shows some important results and properties for linear unconstrained MHE concepts. Especially the connections between the KF, the

Batch estimator and MHE are clarified. Then stability conditions for general initial weights and initial estimates for the linear unconstrained MHE are derived. These results partly coincide with dual results found for the finite horizon linear quadratic regulator. Simulations and examples are presented to clarify that the tuning of the resulting algorithms is not always intuitive. This is similar to the finite horizon LQR or the MPC without final zero endpoint constraint. Additionally some preliminary results about possible cost-functions are presented, which opens the possibility to derive more general stability results for constrained estimation problems.

As a final contribution, a toolbox allowing the easy simulation and inclusion of MHE in closed loop simulations is provided. All of the examples presented in this work where generated using this toolbox. This toolbox allows one to examine conveniently new proposed MHE algorithms and concepts.

Contents

List of Tables

List of Figures

Notation

Upper case symbols

A, B	State transition and input distribution matrices for linear systems.
C, D	Output matrices for linear systems.
$\tilde{A}$	Special matrix section 3.1.4.
B_r	$:= \{x : \|x\| \leq r\},\ r \in \mathbb{R}^+$. Ball with radius r around origin.
$\mathcal{F}$	Linearization of f for EKF, $\mathcal{F}_k = \frac{\partial f(x_k,u_k)}{\partial x_k}\|_{x_k=\hat{x}_{k\|k}}$
G	$\in \mathbb{R}^{n\times r}$ state disturbance distribution matrix.
$\mathcal{G}$	Linearization of g for EKF, $\mathcal{G}_k = \frac{\partial g(x_k)}{\partial x_k}\|_{x_k=\hat{x}_{k\|k-1}}$.
H_x, H_v, H_w	Constraint function or matrixes for x, v, w, e.g. $h_{x\min} \leq H\hat{x} \leq h_{x\max}$ or $h_{x\min} \leq H(\hat{x}) \leq h_{x\max}$.
$\mathcal{K}^0$	A function $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ belongs to the class $\mathcal{K}^0$ if it is (1) continuous and (2) $\alpha(x) = 0 \Leftrightarrow x = 0$.
$\mathcal{K}^+$	α belongs to $\mathcal{K}^+$ if (1) $\alpha \in \mathcal{K}^0$ and (2) it is nondecreasing.
$\mathcal{K}^\infty$	α is in the class $\mathcal{K}^\infty$ if (1) $\alpha \in \mathcal{K}^+$ and (2) $\alpha(x) \to \infty$ when $x \to \infty$.
$\mathcal{K}$	α is called a $\mathcal{K}$-function if (1) $\alpha \in \mathcal{K}^+$ and (2) it is strictly increasing.
K	Gain/filter matrix or nonnegative integer in constrained asymptotic stability definition.
K_∞	Steady state gain for infinite LQR.
L	Linear filter matrix, e.g. KF.
$\mathcal{L}$	nonlinear state dependent penalty for nonlinear GMHE cost function.
M	$\in \mathbb{N}^+$ Positive integer used in Lyapunov theorem, first moment, mean
N	Estimation horizon, discrete time.
$N(a, b)$	Normal distribution with mean a and standard deviation b.

P	Covariance matrix KF, solution of riccati equation.
Q, R	Weighting matrices MHE.
$\tilde{Q}, \tilde{R}$	Weighting matrices MPC.
$\mathcal{Q}, \mathcal{R}$	Weighting matrices GMHE.
$\widetilde{QC}$	Special matrix section 3.1.4.
ΔT	Sampling time.
V	Lyapunov function.
$\hat{\mathcal{V}}$	Constrained output disturbance set.
$\hat{\mathcal{W}}$	Constrained state disturbance set.
$\hat{\mathcal{X}}$	Constrained state estimation set.
X	Whole state vector $X = \{x_{j\|k}\},\ j = 0, \ldots, k$, section 2.3.2.
Y	Whole measurement vector $Y = \{y_{j\|k}\},\ j = 0, \ldots, k$, section 2.3.2.

Lower case symbols

a	$\in \mathbb{R}^+$. Positive constant used in exponential stability definition.
b	Special vector $b^T = \{\bar{x}_{k-N}^T, 0, \vdots, 0, 0\}$, section 3.1.4.
e	Estimation error $e = x - \hat{x}$, $e_{j\|k} = x_j - \hat{x}_{j\|k}$
f	State transition function.
$\tilde{f}^{-1}$	Pseudo inverse of f, section 2.3.2.
g	Output function in dynamic system equations.
$\tilde{g}^{-1}$	Pseudo inverse of g, section 2.3.2.
$h_{x\mathrm{min}}$, $h_{x}\mathrm{max}$	Lower and upper bounds on state, e.g. $h_{x\mathrm{min}} \leq H\hat{x} \leq h_x\mathrm{max}$.
$h_{v\mathrm{min}}$, $h_{v}\mathrm{max}$	Same as above for v.
$h_{w\mathrm{min}}$, $h_{w}\mathrm{max}$	Same as above for w.
j_1, j_2	Time indices.
m	Number of system inputs.
n	Number of system states.
p	Number of system outputs, probability density function.
r	Number of state disturbances, radius $r \in \mathbb{R}^+$.
$p(\cdot\|\cdot)$	Conditional density.

$p(\cdot, \dots, \cdot)$	Joint density.
t	Continuous valued time variable.
u	$\in \mathbb{R}^m$. Input, manipulated variable.
v	$\in \mathbb{R}^p$. Output disturbance variable.
w	$\in \mathbb{R}^r$. State disturbance variable.
$\tilde{\hat{w}}$	Transformed variable $\hat{w}$, section 3.1.4.
x	$\in \mathbb{R}^n$. State space variable.
x^e	Equilibrium or steady state, defined by $x^e := f(x^e)$.
x_0	Initial value.
y	$\in \mathbb{R}^p$. Output, measured variable.
z	Integrated white noise, section 2.4.3.

Greek symbols

α, β	$\mathcal{K}^+$ functions used in Lyapunov theorem.
$\gamma, \delta, \sigma, \epsilon$	$\in \mathbb{R}^+$. Positive real numbers used in stability definitions.
γ	$\mathcal{K}^+$ function
λ	Exponential factor used in exponential stability definition, lagrange multiplier, eigenvalues.
Λ	Matrix used in theorem 3.3 and theorem 3.4.
Π	Riccati matrix MPC.
Π_∞	Steady state riccati matrix MPC.
Φ	Cost function MPC.
Ψ	Cost function MHE.
Υ	Riccati matrix for recursive MHE solution.

Other symbols

$\bar{(\cdot)}$	Denotes continuous time variables and symbols.
$(\cdot)^C$	Denotes colored noise variables.
$(\cdot)^e$	Equilibrium values.
$(\cdot)^P$	Identifier for plant.
$(\cdot)^\star$	Denotes variables or sequences with optimal value.

$(\cdot)'$	Augmented variables or modified regions.
$(\cdot)^T$	Trasnsposed of matrix or vector.
$(\hat{\cdot})$	Estimated value of variable.
$(\cdot)_k$	Discrete time variable at time k.
$(\cdot)_{j\|k}$	Discrete time variable at time j given information up to time k.
$(\cdot)(t)$	Continous time variable at time t.
$(\bar{\cdot})_k$	Value continous time variable during time interval $[t_0 + k\Delta T, t_0 + (k+1)\Delta T]$
$(\cdot)_G$	Combined vectors of section 2.4.3.
$:=$	Left hand side is defined as.
$=:$	Right hand side is defined as.
$\|\|x\|\|$	$:= \sqrt{x^T x}$ Euclidean or l_2 norm of a vector $x \in \mathbb{R}^n$.
$\|\|x\|\|_P$	$:= \sqrt{x^T P x}$ weighted Euclidean norm of a vector $x \in \mathbb{R}^n$ with respect to a positive definite matrix $P \in \mathbb{R}^{n\times n}$
$\|\|A\|\|$	$:= \sup_{x\neq 0, x\in\mathbb{R}^n} \frac{\|\|Ax\|\|}{\|\|x\|\|}$. Induced Euclidean matrix norm of a matrix $A \in \mathbb{R}^{m\times n}$
$\mathbb{C}^j$	Denotes the function space of the j times continuously differentiable functions.
$\mathbb{N}$	The set of nonnegative integers.
$\mathbb{N}^+$	The set of positive integers.
$\mathbb{R}^n$	Euclidean space of dimension n.
$\mathbb{R}^{n\times m}$	Matrix of size $n \times m$.
$\mathbb{R}^+$	Denotes the nonnegative reals.

Chapter 1

Introduction

1.1 Introduction and Motivation

Most control schemes are derived under the assumption that the state of the system is known explicitly. However, the state is often not measured directly in real world applications, so estimates of the actual state based on output measurements must be used instead.

The standard state estimation method used for linear systems is the Kalman filter [18]. The importance of Kalman's approach follows from the fact that it provides a recursive and probabilistic "optimal" solution for *dynamical systems*. Kalman's approach differed from Wiener's method [51] because it can handle dynamic systems, thereby opening the field to a broader class of problems. In addition, given the computational limits of the 1960's, the recursive structure of the Kalman filter provided a tractable solution method.

One drawback of this approach is that a direct transfer to the nonlinear state estimation problem is not possible. As Kushner [25] notes, the exact solution in a probabilistic sense is infinite-dimensional and therefore some approximations must be made. Two examples of approximate methods are the extended Kalman filter (EKF) and statistically linearized filters. These methods base on a linearization around the last estimate. However, these methods can lead to significant deviations between the real state and the estimated state for highly nonlinear systems.

For linear systems it can be shown that the Kalman filter is equivalent to the batch least-squares estimator (BLSE). The BLSE estimates the state by minimizing the squared error between a system model and the output measurements. This method becomes computationally infeasible with increasing time. All available output measurements are used for the estimation. That leads to a steadily growing optimization problem. This computational growth makes the method unusable for real-time problems. It is often used for off-line state and parameter estimation since an expansion to nonlinear systems is possible and the resulting estimates are fairly accurate.

There are several factors that justify the reconsideration of least–squares methods for real–time applications. First of all, the computational limits of the 1960's which suggested a purely recursive formulation have nearly vanished due to the steady advance of computer technologies and numerical optimization methods. Sequential Quadratic Programming (SQP), for example, has greatly reduced the time needed to solve general nonlinear programming problems. Furthermore, the use of optimization based algorithms allows inclusion of additional information about the system in the form of equality and inequality constraints. Bounds on the measurement noise and system states are examples of these constraints[1]. Such information cannot be incorporated in recursive estimation methods and would therefore be lost. Additionally the use of LSE techniques allows the explicit use of nonlinear system models, in contrast to approximation methods like the EKF. There is no need to provide additional information about derivatives, since no linear approximations are necessary.

Motivated by optimization-based control methods such as model predictive control (MPC), that uses a fixed prediction horizon to calculate the next control move, several moving horizon state estimator (MHE) methods have been proposed. These methods partly overcome the problems associated with the growing optimization problem in the batch estimator.

This thesis reviews the necessary filtering and estimation techniques to provide a deeper understanding of the connection between MHE and MPC techniques. Also new results concerning the stability of the linear MHE are presented. Finally a simulation and design tool for MHE methods in conjunction with MPC techniques is provided. This tool facilitates the simulation of nonlinear MPC controllers in connection with general nonlinear MHE based estimation methods.

1.2 Thesis Overview

This work focuses on moving horizon state estimation methods. It is organized in three chapters: a review chapter, a chapter that explores the moving horizon state estimation problem, and a chapter that summarizes the thesis. The chapters are arranged as follows:

- The remainder of Chapter 1 contains some comments about the notation used in this work. The need to distinguish between estimated and real values of variables necessitates the use of a special notation. Most of the work considers

[1]This is of special importance in the case of unreliable or sometimes faulty measurements that would lead to "spikes" in the resulting data.

discrete time systems. An extension to continuous time models is often possible but tedious. The last section of chapter 1 contains some useful definitions and theorems.

- Chapter 2 provides an overview on existing filtering and estimation techniques. A general moving horizon estimation formulation is presented. This method can be motivated from a deterministic or from a probabilistic point of view. The majority of the presented methods can be seen as a special version of the general moving horizon estimation framework. The estimation problem for linear and nonlinear systems is considered separately. This is necessary since nonlinear moving horizon state estimation is still in its infancy and for many algorithms no clear stability proofs exist. The review part for linear systems considers the Kalman filter, the linear MHE, and the Luenberger observer for linear systems. Finally nonlinear state estimation methods are presented. Of special interest are extended Kalman filter methods and nonlinear moving horizon techniques.

- Chapter 3 presents some theoretical results for MHE. First the connections between the linear estimation methods of chapter 2 are examined. Of special interest are the connections between the Kalman filter (KF), the batch state estimator (BSE) and the moving horizon estimator (MHE) with Kalman filter-like update. A stability proof for the batch state estimator is given. Then sufficient conditions for the stability of the MHE formulation with prediction update are presented.
  Additionally some of the properties of the MHE techniques are examined. Different example systems are considered and the estimation problem for these systems is solved using MHE methods. After this the use of Lyapunov stability methods to proof stability of the MHE techniques is motivated. As an example for this the proof of stability for MHE techniques with smoothed update is presented.

- The final chapter serves as a review of the presented work and also discusses future research directions.

1.3 Comments about the Notation

This work is concerned with estimation methods for discrete time systems. Since purely discrete time models cannot be derived for most real systems, continuous time models need also be considered. The continuous-time models must be numerically integrated with online methods to receive the necessary discrete time predictions for the optimization methods. For simplicity, similar symbols for the continuous time and discrete time model variables are used. Continuous time variables are differentiated from discrete time ones by a bar $\bar{(\cdot)}$ (with $(\cdot)$ any valid variable name is meant) . The time index for discrete time variables is given as a subscript, whereas for continuous time variables, the "standard" notation $\cdot(t)$ is used. For example $\bar{x}(t)$ represents the state of a continuous time system at time t, x_k denotes the state of a discrete/discretized system at time element k.

The integer valued time variables correspond to uniform sampling times. The corresponding continuous time points can be calculated using the following formulation $t_k = t_0 + k\Delta T, k \in \mathbb{N}$. Here ΔT denotes the sampling time and t_0 the startup or initial time.

In some cases discrete time indices for continuous time systems are necessary. For example $\bar{u}_k$ mean that the input $\bar{u}(t)$ is kept constant for the time interval spanning from k to $k+1$, $\bar{u}(t) := \bar{u}_k$, $t \in [t_0 + k\Delta T, t_0 + (k+1)\Delta T]$.

In general the estimated state and the real state do not coincide. This makes a distinction necessary. Symbols with an additional circumflex accent $\hat{(\cdot)}$ denote the estimated values. Since optimization based estimation methods using a specific amount of old measurements (sliding window) are used, another distinction is necessary. The double index $(.)_{j|k}$ represents the corresponding variable at time j under the knowledge of the measurement sequence up to time k. For example $\hat{x}_{k-N|k}$ denotes the estimate state for time $k-N$ at time k. Optimal values are marked by a star $(.)^\star$.

Throughout the thesis the symbol ":=" means that the left side is defined to be equal to the right hand side.

1.4 System Models

For simplicity and clarity no time-delays are considered, however if needed they can be incorporated by augmenting the state equation with "memory" stages. Throughout the thesis the following system models are used.

Continuous Time Nonlinear Systems:

Continuous time systems can be given by the following differential equations.

$$\dot{\bar{x}}(t) = \bar{f}(\bar{x}(t), \bar{u}(t), t) + \bar{G}(\bar{x}(t), \bar{u}(t), t)\bar{w}(t) \qquad \bar{x}_0 =: \bar{x}(0) \quad \text{given} \tag{1.1}$$

$$\bar{y}(t) = \bar{g}(\bar{x}(t), \bar{u}(t), t) + \bar{v}(t). \tag{1.2}$$

Often the output is measured only at discrete–sampling time instants

$$\bar{y}_k = \bar{g}_k(\bar{x}_k, \bar{u}_k, k) + v_k. \tag{1.3}$$

Discrete Time Nonlinear Systems:

$$x_{k+1} = f_k(x_k, u_k) + G(x_k, u_k)w_k \qquad x_0 \quad \text{given} \tag{1.4}$$
$$y_k = g_k(x_k, u_k) + v_k. \tag{1.5}$$

Here $k \in \mathbb{N}$ stands for the integer valued time.
The state and output disturbances enter the model equations for both systems linearly. A probabilistic reason for this is given in chapter 2. In both cases $x \in \mathbb{R}^n$ denotes the state, $u \in \mathbb{R}^m$ the input or manipulated variable. Further $w \in \mathbb{R}^r$ represents errors in the state equations, $y \in \mathbb{R}^p$ stands for the output or controlled variable and $v \in \mathbb{R}^p$ denotes measurement errors. $G : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^{n\times r}$ (or $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{N} \to \mathbb{R}^{n\times r}$ in the discrete case) describes the influence of state errors on the system state. If $\bar{f} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n$ (or $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{N} \to \mathbb{R}^n$ in the discrete case) and $\bar{g} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^p$ ($g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{N} \to \mathbb{R}^p$) do not explicitly depend on t (k) then the system is time invariant. In the following g and f are considered time invariant and $\in \mathbb{C}^2$ in order to guarantee local Lipschitz continuity.
Remarks: Often it is not possible to find a closed-form analytical transformation from the continuous time model to the discrete time model. An explicit solution of the state equations with respect to $\bar{u}(t)$ would be necessary. $\triangle$

If f, g are linear in x and u the system is called linear:

$$\dot{\bar{x}}(t) = \bar{A}(t)\bar{x} + \bar{B}(t)\bar{u}(t) + \bar{G}(t)\bar{w}(t) \qquad \bar{x}_0 := \bar{x}(0) \quad \text{given} \tag{1.6}$$
$$\bar{y}(t) = \bar{C}(t)\bar{x} + \bar{D}(t)\bar{u}(t) + \bar{v}(t). \tag{1.7}$$

In the case of discrete measurements we get the corresponding output equation:

$$y = C\bar{x}_k + D\bar{u}_k + \bar{v}_k. \tag{1.8}$$
$$\tag{1.9}$$

Linear discrete time systems can be described by the following state equations:

$$x_{k+1} = A_k x_k + B_k u_k + G_k w_k \qquad x_0 \quad \text{given} \tag{1.10}$$
$$y_k = C_k x_k + D_k u_k + v_k. \tag{1.11}$$

1.5 Mathematical Definitions

During the rest of the thesis $\mathbb{R}^+$ denotes the nonnegative reals.

Definition 1.1 (Positive definite matrices) *We say a matrix $A \in \mathbb{R}^{n\times n}$ is positive definite if $\forall x \neq 0,\ x^T Ax > 0$.*

Definition 1.2 (Matrix Inequalities) *The statement $A > B$ means that the matrix $A - B$ is positive definite.*

Definition 1.3 (Norms) *The symbol $\|x\|$ denotes the Euclidean or l_2 norm of a vector $x \in \mathbb{R}^n$, whereas $\|x\|_P$ denotes the weighted Euclidean norm of a vector with respect to a positive definite matrix $P \in \mathbb{R}^{n\times n}$. In the case of a matrix $A \in \mathbb{R}^{m\times n}$ $\|A\|$ means the induced matrix norm:*

$$\begin{aligned} \|x\| &:= \sqrt{x^T x} \\ \|x\|_P^2 &:= x^T P x \\ \|A\| &:= \sup_{x\neq 0, x\in\mathbb{R}^n} \frac{\|Ax\|}{\|x\|}. \end{aligned}$$

Definition 1.4 (Region $\boldsymbol{B_r}$) *The set B_r with $r \in \mathbb{R}^+$ denotes the ball with radius r in $\mathbb{R}^n$.*

$$B_r := \{x : \|x\| \leq r\}$$

Definition 1.5 ($\boldsymbol{\mathcal{K}^0}$ function) *A function $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ belongs to the class $\mathcal{K}^0$ if (1) it is continuous and (2) $\alpha(x) = 0 \Leftrightarrow x = 0$.*

Definition 1.6 ($\boldsymbol{\mathcal{K}^+}$ function) *α belongs to $\mathcal{K}^+$ if (1) $\alpha \in \mathcal{K}^0$ and (2) it is nondecreasing*

Definition 1.7 ($\boldsymbol{\mathcal{K}^\infty}$ function) *α is in the class $\mathcal{K}^\infty$ if (1) $\alpha \in \mathcal{K}^+$ and (2) $\alpha(x) \to \infty$ when $x \to \infty$*

Definition 1.8 ($\mathcal{K}$-function) *A function α is called a $\mathcal{K}$-function if (1) $\alpha \in \mathcal{K}^+$ and (2) it is strictly increasing.*

Remarks: Notice the difference between a $\mathcal{K}$-function (Definition 1.8) and a function that belongs to the class $\mathcal{K}^+$ (Definition 1.6). A $\mathcal{K}$-function is strictly increasing whereas a function of the class $\mathcal{K}^+$ is only nondecreasing. $\triangle$

Since we deal with constrained systems, it is necessary to modify slightly the usual Lyapunov stability definitions as used in Kalman and Bertram ([19],[20]), Kwakernaak and Sivan [26] or Vidyasagar [50]. The stability definitions and theorems are modified versions of the ones Keerthi and Gilbert [21] presented. We consider time-invariant, nonlinear discrete time systems:

$$x_{k+1} = f(x_k), \qquad x_k \in \mathcal{X} \subset \mathbb{R}^n \tag{1.12}$$

with $f : \mathcal{X} \to \mathcal{X}$, where $\mathcal{X}$ denotes a constrained subspace of $\mathbb{R}^n$.

Definition 1.9 (Steady State for Constrained Systems) *A state $x^e \in \mathbb{R}^n$ is called an equilibrium or steady state for the constrained system (1.12) if (1) $x^e \in \mathcal{X}$ and (2) $x^e = f(x^e)$.*

Definition 1.10 (Constrained Stability) *The steady state $x^e = 0$ is stable if, for every $\epsilon > 0$, $\exists\, \delta(\epsilon) > 0$, such that*

$$\|x_j\| \leq \epsilon, \qquad \forall j \geq k,\ x_k \in \mathcal{X},\ x_k \in B_\delta.$$

Remarks: The major difference between this stability definition and the traditional stability definition is that the considered states must lie not only in the region B_δ but also in the set $\mathcal{X}$. $x_k \in \mathcal{X}$ enforces $x_j \in \mathcal{X}\ \forall j \geq k$ since $f : \mathcal{X} \to \mathcal{X}$. $\triangle$

Definition 1.11 (Constrained Asymptotic Stability) *The origin is asymptotically stable, if (1) it is stable, (2) $\exists\ \gamma > 0$ and, for any $\sigma > 0\ \exists\ K(\sigma) \in \mathbb{N}$ such that*

$$\|x_j(x_k)\| \leq \sigma \qquad \forall j \geq k + K, \quad x_k \in \mathcal{X}, \quad x_k \in B_\gamma.$$

Remarks: Notice that the second requirement implies for all $\|x_k\| \in B_\gamma$ that $\|x_j\| \to 0$ as $j \to \infty$. $\triangle$

Definition 1.12 (Constrained Exponential Stability) *The origin is exponentially stable if there exists $\delta > 0, a > 0$, and $0 \leq \lambda < 1$ such that*

$$\|x_j(x_k)\| \leq a\|x_k\|\lambda^{j-k} \qquad \forall j \geq k,\ \|x_k\| < \delta,\ x_k \in \mathcal{X}.$$

The following theorem is taken from Keerthi [21].

Theorem 1.1 (Lyapunov) *Given a function $V : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following:*
There exist $\alpha, \beta \in \mathcal{K}^+$, $\gamma \in \mathcal{K}^0$, $r \in \mathbb{R}^+$ and a positive integer $M \in \mathbb{N}^+$, such that

1. $V(x) \leq \beta(\|x\|), \quad \forall x \in \mathcal{X},\ x \in B_r$
2. $\alpha(\|x\|) \leq V(x), \quad \forall x \in \mathcal{X}$
3. $V(x_j) - V(x_{j+M}(x_j)) \geq \gamma(\|x_j\|),\ V(x_j) - V(x_{j+1}(x_j)) \geq 0, \quad \forall x_j \in \mathcal{X}$.

Then the zero steady-state solution of (1.12) is locally asymptotically stable with the region of attraction $\mathcal{X}$.

Chapter 2

State Estimation Review

For many real systems the state cannot be measured explicitly. Therefore state estimators (SE) giving an estimate of the actual state, based on measurements, are required to implement state feedback control strategies. Model predictive control techniques in particular require a precise knowledge of the current state of the system in order to solve the the "optimal control" problem.
The two basic goals of state estimation are:

1. Get estimates of unmeasured states from output measurements.

2. Reduce the influence of measurement noise on state estimates.

The combination of SE and state feedback control is shown in figure 2.1.

For linear stochastic systems the Kalman filter [18] provides, the optimal solution to this problem. The process and output noise are assumed to be Gaussian random variables with zero mean. For nonlinear systems such an optimal strategy is in general not available. There exist different proposed methods. It is the control engineer's task to decide upon one method best suited for the considered system.
The remainder of this chapter is subdivided into the following sections:

- Part one contains a presentation of the considered system models under the influence of disturbances.

- The second part gives an overview of existing SE methods and introduces a general framework for optimization based state estimators.

- Part three contains probabilistic and deterministic motivations and interpretations of least squares state estimation methods.

- Part four contains a review of linear SE. The majority of the presented algorithm is of the least squares type, like the batch least squares estimator, the moving horizon estimator with quadratic cost and the Kalman filter. Another

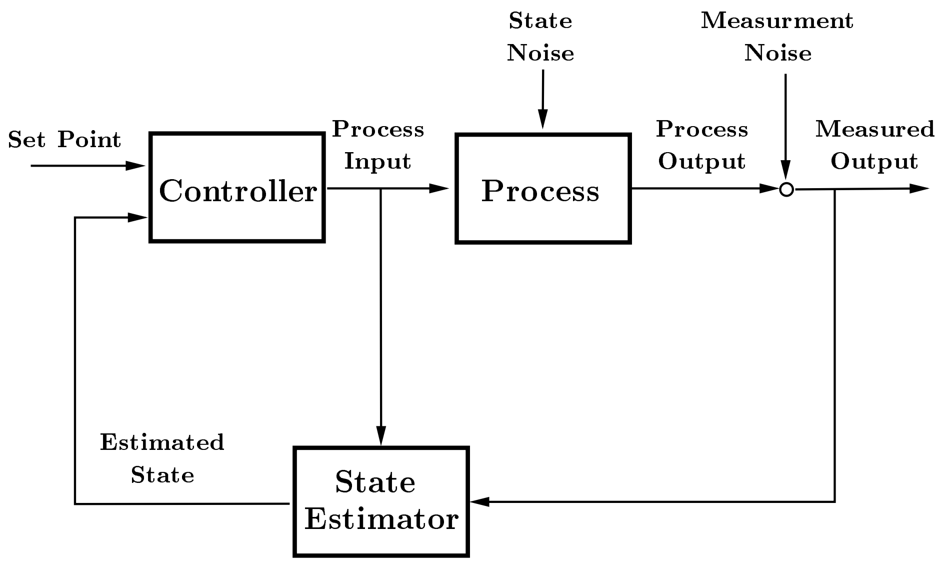

Figure 2.1: SE and state feedback control.

approach for the solution of the linear state estimation problem is the Luenberger observer.

- The last section contains a review of nonlinear state estimation. The presented algorithms are the batch and nonlinear moving horizon state estimator, the extended Kalman filter, the extended Luenberger observer and statistical approximation methods.

2.1 Stochastic System Models

In the following chapters we considered the nonlinear discrete-time invariant system under disturbances:

$$\begin{aligned} x_{k+1} &= F(x_k, u_k) + G w_k, \qquad k = 0, 1, 2, \ldots \\ y_k &= g(x_k, u_k) + v_k \end{aligned} \tag{2.1}$$

$x_k \in \mathbb{R}^n$ is the state of the system, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ describes the propagation of the state x from time k to $k+1$ as a function of the previous state x_k and the input $u_k \in \mathbb{R}^m$. $w_k \in \mathbb{R}^r$ stands for the state noise vector, which enters the system via the distribution matrix $G \in \mathbb{R}^{n \times r}$. w_k can be interpreted as an additional random driving "force" acting on the system. $y_k \in \mathbb{R}^p$ represents the output variable which depends via the nonlinear function $g : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p$ on the state x_k and the input u_k. Another contribution to the output measurement is the measurement noise vector $v_k \in \mathbb{R}^p$.

Remarks: The probabilistic motivation of the least squares estimator clarifies why it is necessary that w_k and v_k enter the system in linear form. For the deterministic motivation given in section 2.3.1 it is necessary that v_k and w_k enter the system decoupled from the input u_k and state x_k. △

Most of this work does not consider the inputs u_k. For linear systems this is not restriction since the system can be shifted by the nominal trajectory resulting from the input u_k. However for nonlinear systems this is not possible. Significant changes of the presented nonlinear algorithms might be necessary if inputs u are considered.

The Meaning of G, a Probabilistic Motivation

From a probabilistic point of view, G offers the possibility to consider colored noise resulting from white noise. For this the system is augmented by additional states w_k^C that represent the colored noise terms.

$$\begin{aligned} \underbrace{\begin{bmatrix} x_{k+1} \\ w_{k+1}^C \end{bmatrix}}_{x'_{k+1}} &= \underbrace{\begin{bmatrix} f(x_k,u_k)+ & B^C w_k^C \\ & A^C w_k^C \end{bmatrix}}_{f'(x'_k,u_k)} + \underbrace{\begin{bmatrix} G \\ G^C \end{bmatrix}}_{G'} w_k \end{aligned} \tag{2.2}$$

$$\begin{aligned} y_k &= \underbrace{g(x_k,u_k)+g^C(w_k^C)}_{g'(x'_k,u_k)} + \underbrace{v_k}_{v_k} \end{aligned}$$

A deterministic motivation for G might be the possibility that the same disturbance can enter different states. One example would be a chemical reactor with the influence of the surrounding temperature modeled as an additional state disturbance. Changes in the surrounding temperature would influence the jacket fluid temperature along with the reactor temperature.

2.2 Overview on State Estimation Methods

The state estimation problem can be considered and tackled from two different points.

The Deterministic Point of View

In the first case the estimation problem is treated as purely deterministic. Mostly this is considered if no satisfactory information about the disturbance v_k and w_k is available. Another reason might be the fact that the influences of v_k and w_k on the system are negligible. The task is then often considered as the optimal recovery from a wrong initial state estimate or as the extraction of the non-measurable states from the outputs.

Most of the algorithms that fall under these category are designed to achieve an "optimal" recovery from an incorrect startup value. The resulting SE methods are called observers. Design methods similar to multivariable control techniques like pole placement are employed to achieve the desired behavior.

Estimation schemes from this group include the Luenberger observer [29] for linear systems, observers based on geometric input-output linearization techniques for scalar systems like the canonical normal form observers [4] and extended Luenberger observers [5, 23]. These methods can be generalized for multivariable systems [24, 52].

The Probabilistic Point of View

The second case uses probabilistic information about w_k and v_k. The resulting estimates are often optimal in the sense that they are most probable under the given output information. Methods based on these ideas are in general called filters.

Stability for these methods is often only shown for the nominal case. This means that a from a wrong initial estimate $\bar{x}_0$ resulting estimate $\hat{x}_k$ under the influence of *no disturbance* will converge to the real state x_k.

In order to apply these techniques, the probability distributions of v_k, x_0, w_k: $p_{v_k}(v_k)$, $p_{x_0}(x_0)$, $p_{w_k}(w_k)$ must be known. Under this information the estimation problem can be seen as the calculation of a "maximum" of the resulting probability density function. For linear systems a recursive solution, the Kalman filter [18] can be derived. An extension to the general nonlinear case via the so called extended Kalman filters [16, 13] is possible.

Other nonlinear state estimation methods are statistical approximation methods. These approximate the system and measurement equations by polynomial series [13, 47].

Most of the subsequently presented optimization based estimation methods can be derived from a probabilistic and deterministic point of view. For the probabilistic derivation, the disturbance is considered to be Gaussian with zero mean with no additional constraints on the states or disturbances. Section 2.3.2 gives a probabilistic motivation of the so called least squares estimation methods for models with the structure as in (2.1).

These methods were explored in the late 1960's [8, 16]. Computational and numerical limitations made a practical implementation impossible at that time. These methods gained interest in recent years [30, 41, 35, 49]. One of the reasons for this is

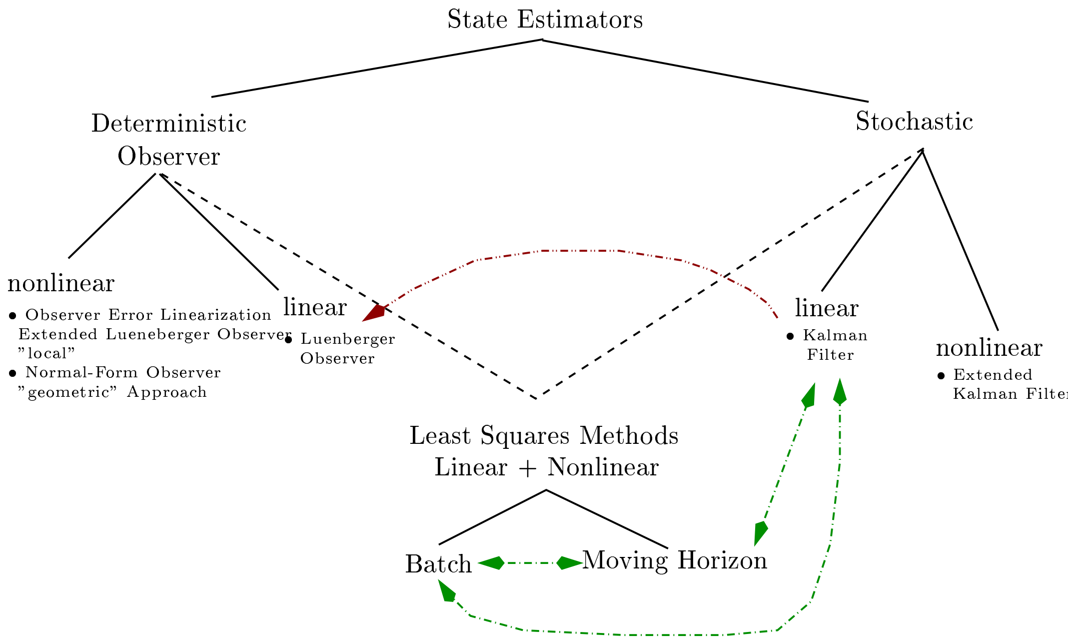

Figure 2.2: Overview of estimation techniques.

the possibility to include explicitly additional information on the disturbances, such as upper and lower bounds. These bounds on the disturbances, however make it impossible to derive the resulting algorithms via a purely probabilistic way, since the underlying distributions are no longer Gaussian. On the other hand it seems to be natural to include bounds, since the real noise is normally not unlimited in size. The resulting methods can outperform standard techniques like the Kalman filter and the extended Kalman filter [41, 35, 49].

A survey of the considered state estimation methods is given in Figure 2.2.

Remarks: It is worth pointing out that for linear systems, with no additional constraints, connections between the probabilistic and deterministic approaches can be derived. One example is the equivalence of the moving horizon estimator, the Kalman filter and the batch state estimator shown in chapter 3. △

2.3 A General Moving Horizon Estimator Formulation GMHE

Most of the methods presented in the following sections are based on a least squares estimation scheme. A general version of this scheme, in the following general moving horizon estimator (GMHE) can be given as follows:

$$\min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k : \quad \Psi_k = \hat{w}^T_{k-N-1|k}\mathcal{Q}_{-N|k}\hat{w}_{k-N-1|k} + \sum_{j=k-N}^{k-1} \hat{w}^T_{j|k}\mathcal{Q}\hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}^T_{j|k}\mathcal{R}\hat{v}_{j|k} \quad (2.3)$$

subject to the state equality constraints

$$\begin{aligned} \hat{x}_{k-N|k} &= \bar{x}_{k-N} + G\hat{w}_{k-N-1|k} \\ \hat{x}_{j+1|k} &= f(\hat{x}_{j|k}) + G\hat{w}_{j|k} \qquad , j = k-N\dots k-1 \\ y_j &= g(\hat{x}_{j|k}) + \hat{v}_{j|k} \qquad , j = k-N\dots k \end{aligned} \quad (2.4)$$

Here $\mathcal{R} = \mathcal{R}^T > 0 \in \mathbb{R}^{p\times p}$, $\mathcal{Q} = \mathcal{Q}^T > 0 \in \mathbb{R}^{r\times r}$, $\mathcal{Q}_{-N|k} = \mathcal{Q}^T_{-N|k} \in \mathbb{R}^{r\times r} > 0$ and $\bar{x}_{k-N}$ an estimate of the "initial" state at time $k-N$. $N \in \mathbb{N}^+$ represents the estimation horizon length or window, see Figure 2.3.

$\mathcal{R}$ penalizes the output prediction error, $\mathcal{Q}$ penalizes the estimated noise vector and $\mathcal{Q}_{-N|k}$ penalizes the initial estimation error. If the expected output error is small, than $\mathcal{R}$ is chosen large compared to $\mathcal{Q}$. The $\hat{v}_{j|k}$'s resulting from the optimization will be small compared to the corresponding $\hat{w}_{j|k}$'s to keep the cost low. In the case of unreliable output measurements $\mathcal{Q}$ is chosen large compared to $\mathcal{R}$.

Additional bounds on the states $x_{j|k}$, inputs $u_{j|k}$ and the disturbances $w_{j|k}$ and $v_{j|k}$ might be considered.

This GMHE concept can be motivated under additional assumptions on the state and output noise from a probabilistic point of view. However a purely deterministic motivation is also possible. This is especially important from an engineering point of view, since it allows the easy addition of constraints on the disturbances without having to worry about the introduced changes in the probability density functions.

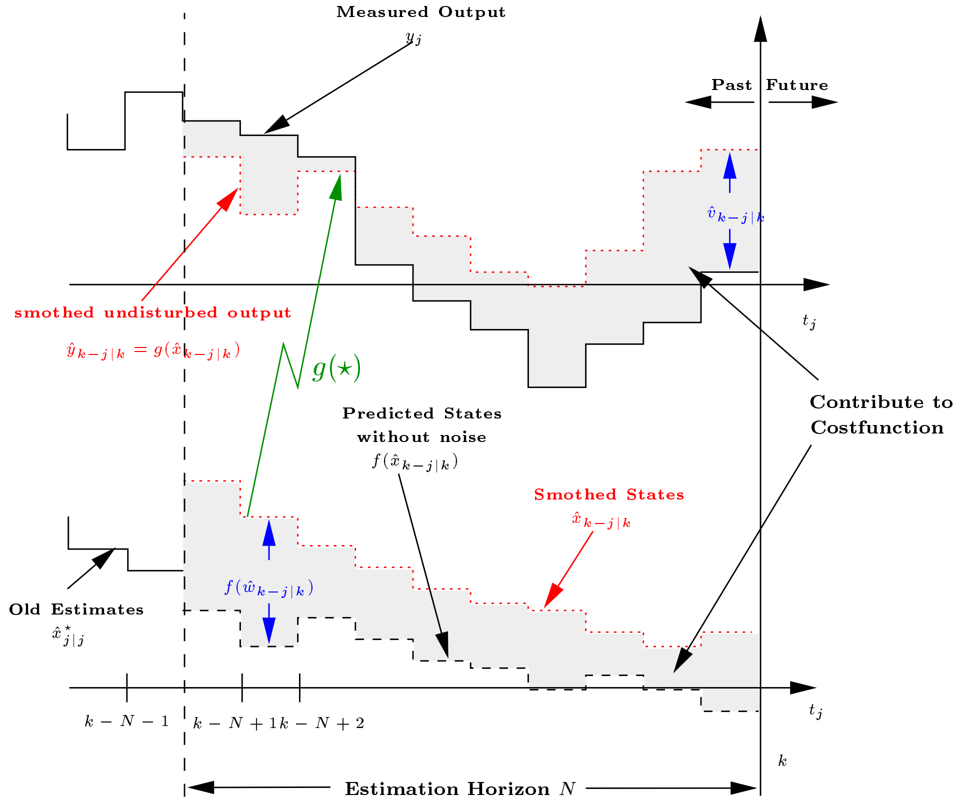

Figure 2.3: General moving horizon state estimation and the cost function.

2.3.1 A Deterministic Motivation for the GMHE

For a deterministic motivation of the GMHE consider the following vector difference equations.

$$\begin{aligned} x_{j+1} &= f(x_j) + G(x_j)w_j, \qquad j = 0, 1, 2, \dots \\ y_j &= g(x_j) + v_j \end{aligned} \tag{2.5}$$

Here x_j, u_j, w_j and v_j have the same dimensions as above. $G : \mathbb{R}^n \to \mathbb{R}^{n \times r}$ describes the influence of w_j depending on the actual state of the system. Following a presentation given in [16] w_j and v_j are not considered as disturbances with well defined statistics. Instead they represent errors of unknown character, so that (2.5) is not a stochastic difference equation. Rather it is as an ordinary difference equation that could be solved if the errors were known. w_j is a deterministic modeling error or an un modeled disturbance term. Suppose that a noisy sequence of observations y_j up to time k is available. Under the additional assumption that a prior estimate of x_0, called $\bar{x}_0$ is available, the sequence $\{x_0, \dots, x_k\}$ is calculated (estimated) so that the errors w_k and v_k in (2.5) are minimized. A practical interpretation of this might be that the solution of (2.5) should be passed as closely as possible through the observations $\{y_0, \dots, y_k\}$.

One possible method to achieve this might be the classical least squares approach. For this the following function must be minimized with respect to the $w_{j|k}$ sequence:

$$\Psi_k = \hat{w}_{-1|k}^T \mathcal{Q}_0 \hat{w}_{-1|k} + \sum_{j=0}^{k-1} \hat{w}_{j|k}^T \mathcal{Q} \hat{w}_{j|k} + \sum_{j=0}^{k} \hat{v}_{j|k}^T \mathcal{R} \hat{v}_{j|k} \tag{2.6}$$

subject to the equality constraints:

$$\begin{aligned} \hat{x}_{0|k} &= \bar{x}_0 + \hat{w}_{-1|k} \\ \hat{x}_{j+1|k} &= f(\hat{x}_{j|k}) + G(\hat{x}_{j|k})\hat{w}_{j|k}, \quad j = 0, 1, 2, \dots, k-1 \\ y_j &= g(\hat{x}_{j|k}) + \hat{v}_{j|k}, \end{aligned} \tag{2.7}$$

Here the circumflex accent $\hat{\cdot}$ and the double index $\cdot_{j|k}$ are used, since the optimization values do not coincide with the "real" values of equation (2.5).

The first term corresponds to our belief in the given initial estimate $\bar{x}_0$. $\mathcal{R}$, $\mathcal{Q}$ and $\mathcal{Q}_0$ are symmetric positive definite and can be seen as weighting matrices. They are quantitative measurements of the belief in the observation equation, the dynamical part of the system model and the prior estimate.

The estimation problem as stated in equation (2.6) is a batch optimization problem. The estimate of x_k depends on all of the information y_j up to this time. This would lead to a steadily growing number of optimization variables as time increases. There are different ways to overcome this problem. In the unconstrained linear case it is possible to solve the batch optimization recursively. This leads to a deterministically derived version of the Kalman filter. Another way would be to consider separate batches. Batch one would only use information from $k = 0$ to $k = j_1$. The next batch would consider measurements y_j from $k = j_1 + 1$ to $k = j_2$ and so on. The resulting algorithm is often referred to as batch processing. This is not a good idea since no information is carried over from calculation to calculation. A different approach is the GMHE introduced in (2.3). In this case a "window" of the size N is moved over the output measurements, see figure 2.3. Information from previous calculations can be carried over with the "initial" estimate $\bar{x}_{k-N|k}$ and $\mathcal{Q}_{-N|k}$.

The major advantage of this deterministic least squares motivation is the possibility to include constraints on the errors with out having to modify the considered distributions. This allows us to explicitly include information about the errors, like maximum and minimum values.

2.3.2 A Probabilistic Motivation for the GMHE

LSE methods also arise from a probabilistic formulation of the SE problem. SE in a probabilistic framework can be summarized as follows:

> Given the start up estimate, current and past measurements $Y = \{y_{j|k}\},\ j = 0, \ldots, k$ and their probability distributions, find the probability distribution of the state $X = \{x_{j|k}\}$ and extract from this the most probable state estimate $\hat{X} = \{\hat{x}_{j|k}\}$.

A closed-form analytic expression of the probability distribution of the states is very difficult to derive, even if the considered noise and startup estimates are Gaussian with zero mean. However a closed analytic solution, is often not necessary. The calculation of the state estimate can be seen as a search for an "optima" of the probability density function with respect to an adequate norm. Solutions to this problem are possible without full knowledge of the whole probability density function.

Criteria often used to compute estimates $\hat{X} = \{\hat{x}_{j|k}\}$ from $p(X|Y)$ are:

- Criterion: Maximize the probability density
  Solution: $\hat{X} =$ mode/maxima of $\quad p(X|Y)$
  This can be seen as the most probable "single estimate". When the given prior noise and state density functions are uniform this estimate is the maximum Liklehood estimate.

- Criterion: Minimize $\int \|X - \hat{X}\|^2 \, p(X|Y) \, dX$
  Solution: $\hat{X} = E(X|Y)$.
  This is the conditional mean.

- Criterion: Minimize the maximum of $|X - \hat{X}|$
  Solution: $\hat{X} =$ Median of $p(X|Y)$.
  This is the so called median or min-max estimate

A visualization for the scalar case of these criteria is given in figure 2.4 . Since a closed analytical form of $p(X|Y)$ is not available, the usage of the conditional mean and the min-max estimate is not favorable. Therefore the mode is instead used often as the best estimate for X.

The following system model is considered:

$$\begin{aligned} x_{k+1} &= f(x_k, w_k), \qquad k = 0, 1, 2, \ldots \\ y_k &= g(x_k, v_k) \end{aligned} \tag{2.8}$$

This system model is different from that considered in (2.1).The following derivation clarifies, why it makes sense to consider the simpler model.

Additional assumptions are:

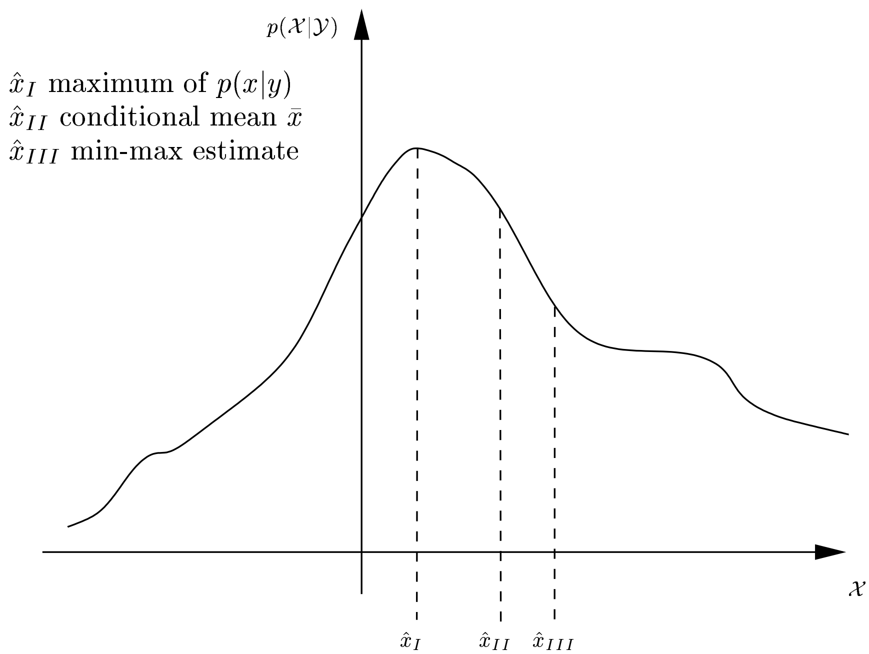

Figure 2.4: Different probabilistic estimation criteria.

A1 The distributions for v_k, x_0, w_k: $p_{v_k}(v_k)$, $p_{\bar{x}_0}(\bar{x}_0)$, $p_{w_k}(w_k)$ are known.

A2 $f, g \in C^1$

A3 $p_{v_k}(v_k)$,$p_{\bar{x}_0}(\bar{x}_0)$,$p_{w_k}(w_k)$ are independent uncorrelated random variables.

The calculation of the estimate that maximizes the probability density function can be divided into 2 steps:

Step 1. derivation of the probability of all $\{x_{j|k}\} = X_k$ under the the known measurement sequence $\{y_j\} = Y_k$.

Step 2. calculation of the mode of $p(x_k, x_{k-1}, \ldots, x_0|y_k, y_{k-1}, \ldots, y_0)$.

Step1:
Expressing $p(x_k, x_{k-1}, \ldots, x_0|y_k, y_{k-1}, \ldots, y_0)$ using Bayes theorem results in:

$$p(x_k, x_{k-1}, \ldots, x_0|y_k, y_{k-1}, \ldots, y_0) = \frac{p(y_k, y_{k-1}, \ldots, y_0|x_k, x_{k-1}, \ldots, x_0)p(x_k, x_{k-1}, \ldots, x_0)}{p(y_k, y_{k-1}, \ldots, y_0)} \tag{2.9}$$

From equation (2.8) for x_{k+1} it can be seen that each x_{k+1} depends only on the previous x_k and $p_{w_k}(w_k)$. This implies that the x_k sequence is a Markov Chain.

As a consequence of this property the y_k sequence is also Markov. Using now the "decoupling" Markov property leads to:

$$\begin{aligned} p(y_k, y_{k-1}, \dots, y_0 | x_k, x_{k-1}, \dots, x_0) &= \prod_{i=0}^{k} p(y_i|x_i) \\ p(x_k, x_{k-1}, \dots, x_0) &= \left(\prod_{i=1}^{k} p(x_i|x_{i-1}) \right) p(x_0) \end{aligned}$$

Use of the probability transfer function theorem as given in appendix A.1 for the calculation of the new probabilities results in

$$\begin{aligned} p(y_k, y_{k-1}, \dots, y_0 | x_k, x_{k-1}, \dots, x_0) &= \prod_{i=0}^{k} p_{v_k} \left(\tilde{g}^{-1}(y_i, x_i) \left\| \frac{\partial \tilde{g}^{-1}(y_i, x_i)}{\partial y_i} \right\| \right) \\ p(x_k, x_{k-1}, \dots, x_0) &= (\prod_{i=1}^{k} p_{w_k} \left(\tilde{f}^{-1}(x_i, x_{i-1}) \left\| \frac{\partial \tilde{f}^{-1}(x_i, x_{i-1})}{\partial x_i} \right\| \right) p(x_0) \end{aligned} \tag{2.10}$$

Remarks: The result as given in (2.10) is achieved by considering $p(y_j|x_j) = p(y_j)$. For the calculation the premise variables x_j are fixed, which means that they are not random anymore. $\tilde{g}^{-1}(y_i, x_i) = v_i$ is the inverse of $g(x_i, v_i)$ with respect to v_i and $x_i = \text{const}$. This is not the general inverse of g with respect to x_{i-1}, w_i. The same holds for $\tilde{f}^{-1}(x_i, x_{i-1}) = w_{i-1}$ as the inverse of $f(x_{i-1}, w_{i-1})$ with respect to w_{i-1}. △

To reformulate (2.9) an additional equation for $p(y_k, y_{k-1}, \dots, y_0)$ is needed. This density depends only on the y_k sequence and therefore does not change with the choice of x_k. By this it can be seen as a constant with respect to $\{x_k\}$. As a result it is easy to see that this value has no influence on the location of the maxima of $p(x_k, x_{k-1}, \dots, x_0 | y_k, y_{k-1}, \dots, y_0)$ for a given $\{y_k\}$ sequence.

Step 2:

Using the relations derived in step 1, the following optimization problem has to be solved for the calculation of the mode $\hat{X}_k^\star$:

$$\begin{aligned} &\arg \max_{\{x_k, x_{k-1}, \dots, x_0\}} p_{v_k} \left(\tilde{g}^{-1}(y_i, x_i) \left\| \frac{\partial \tilde{g}^{-1}(y_i, x_i)}{\partial y_i} \right\| \right) \cdot \\ &\left(\prod_{i=1}^{k} p_{w_k} \left(\tilde{f}^{-1}(x_i, x_{i-1}) \left\| \frac{\partial \tilde{f}^{-1}(x_i, x_{i-1})}{\partial x_i} \right\| \right) \right) p(x_0) \end{aligned} \tag{2.11}$$

This is the general nonlinear probabilistic mode estimate for $\hat{X}_k^\star$. To derive the deterministic least squares approach from here the following additional restrictions are necessary

A4 w_k and v_k enter the system in the following linear, decoupled form[1]:

$$\begin{aligned} x_{k+1} &= f(x_k) + w_k, \qquad k = 0, 1, 2, \ldots \\ y_k &= g(x_k) + v_k \end{aligned}$$

A5 All random variables are Gaussian normal distributed:

$$w_k \sim N(0, Q), \quad v_k \sim N(0, R), \quad \bar{x}_0 \sim N(\bar{x}_0^M, Q_0)$$

Under these conditions $\tilde{f}^{-1}(x_k, x_{k-1}) = x_{k+1} - f(x_k)$,$\tilde{g}^{-1}(y_i, x_i) = y_k - g(x_k)$ and the resulting optimization is a maximization of an exponential function with negative exponent. This allows to simplify (2.11) to the following minimization problem:

$$\hat{X}_k^\star = \arg \min_{\{x_k, x_{k-1}, \ldots, x_0\}} \|x_0 - \bar{x}_0\|^2_{Q_0^{-1}} + \sum_{i=1}^{k} \|x_i - f(x_{i-1})\|^2_{Q^{-1}} + \sum_{i=0}^{k} \|y_i - g(x_i)\|^2_{R^{-1}} \quad (2.12)$$

$\hat{X}_k^\star$ becomes $\hat{X}_k^\star\{x_k^\star, x_{k-1}^\star, \ldots, x_0^\star\}$. This equals the batch GMHE least squares approach with the inverse covariance matrices R^{-1}, Q^{-1}, Q_0^{-1} as the weighting matrices $\mathcal{R}$, $\mathcal{Q}$ and $\mathcal{Q}_0$. From this it follows that the least squares formulation of the estimation problem arises naturally from a probabilistic setup.

Figure 2.3 gives a graphical interpretation of the optimization problem and the connections between the measurements, estimates and real states for the GMHE. In this picture an expansion for a moving horizon "estimation" window is already made. For the equations derived here N equals k, which corresponds to a steadily growing filtering horizon. The resulting least squares estimator in conjunction with a state feedback controller is given in figure 2.5.

2.3.3 Remarks about the GMHE

As shown in section 2.3.2 and 2.3.1 the use of a least squares approach for optimization based SE is natural. However it is possible to use a nonlinear state depending cost function instead of a quadratic one in the optimization.

$$\Psi_k = \mathcal{L}_{-N|k}(w_{-1|k}, \hat{x}_{k-N|k}) + \sum_{j=k-N}^{k-1} \mathcal{L}(w_{j|k}, v_{j|k}, \hat{x}_{j|k}) + \mathcal{L}_{final}(v_{j|k}, \hat{x}_{k|k}) \quad (2.13)$$

[1]For a more general and rigorous treatment, even for noise entering the system as given in a form similar to (2.5) see [12]

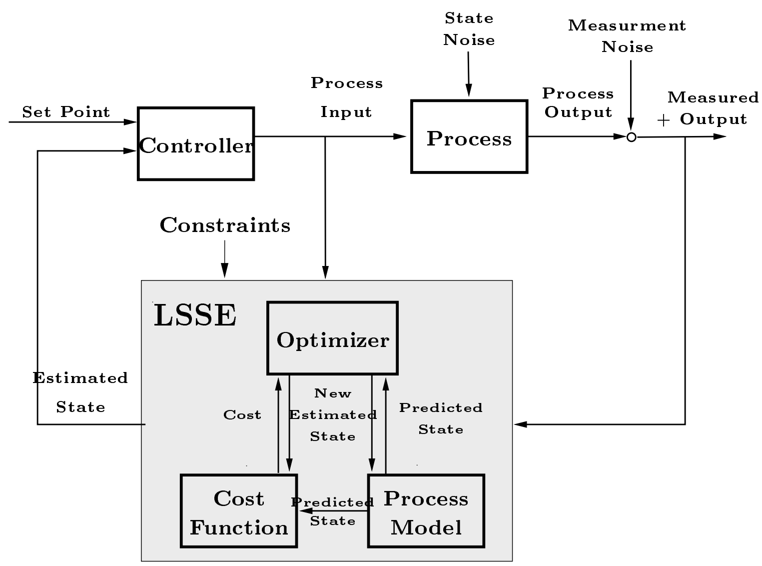

Figure 2.5: Basic structure of optimization based SE.

This approach is however not considered here since the whole area of moving horizon state estimation is still in an infant state. An understanding of the normal least squares approach is appropriate before considering the more general formulations.

2.4 Linear State Estimation

This section discusses the application of least square based methods to linear systems. For linear systems it is possible to derive a recursive solution algorithm for the least squares method that allows the faster calculation of the next state estimate in dependence of the last one calculated. This recursive formulation was first derived by Kalman [18]. Kalman's state space method renewed intense interest in filtering theory by overcoming the restrictive stationarity assumption of the Wiener-Kolmogorov theory of linear filtering, opening new perspectives. For linear systems with Gaussian noise structure which are considered in the Kalman filter formulation, the decision which maxima of the probability density function to take is easy, since the mode and the conditional mean are the same. Therefore there is no reason for using a non-recursive formulation in order to calculate the next estimate. However the power of the batch formulation lies in the possibility to incorporate additional constraints describing prior system knowledge like the restriction of the state to a certain region or physical limitations of the influencing noises. This addition of constraints is not possible for the recursive formulations. Important information about specific system characteristics is lost which could help to improved or faster convergence of the esti-

mate.
First the batch state estimator is presented, after this the Kalman filter as a recursive solution of the unconstrained batch estimator is discussed. Following this presentation the moving horizon estimators are motivated and two different approaches guaranteeing stability under constraints are discussed. The following part gives a short introduction of the basic ideas of linear observer design methods. The end of the section points out that all linear estimator methods can be reduced to the same basic structure.
The system models considered in this section are linear and have the following time-invariant structure similar to the structure given in (1.10) and (1.11)

$$\begin{aligned} x_{k+1} &= Ax + Bu_k + Gw_k \qquad x_0 \quad \text{given} && (2.14)\\ y_k &= Cx_k + Du_k + v_k && (2.15) \end{aligned}$$

2.4.1 Batch Estimator

If the LS formulation resulting from the probabilistic/deterministic derivation in 2.3.1, 2.3.2 is used as a method to estimate the states of a system, we talk about the so called batch least squares estimator. This estimator can also be seen from the GMHE viewpoint with N equal to k and therefore time-variant.

$$\begin{aligned} \min_{\{\hat{w}_{-1|k},\dots,\hat{w}_{k-1|k}\}} \Psi_k &= \hat{w}_{-1|k}^T Q_0^{-1} \hat{w}_{-1|k} + \sum_{j=0}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} \qquad (2.16)\\ &+ \sum_{j=0}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \end{aligned}$$

$$\text{Subject to:} \quad \begin{aligned} \hat{x}_{0|k} &= \bar{x}_0 + \hat{w}_{-1|k}\\ \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + Bu_j + \hat{w}_{j|k}\\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k} \end{aligned} \qquad (2.17)$$

The state estimates resulting from this minimization can be calculated using the following equation in dependence of the calculated $\{\hat{w}^{\star}_{j|k}\}$ sequence:

$$\hat{x}^{\star}_{i|k} = A^i \bar{x}_0 + \sum_{j=0}^{i} A^{i-j} \hat{w}^{\star}_{j-1|k} + \sum_{j=0}^{i} A^{i-j-1} Bu_j$$

This formulation has the problem that the optimization problem grows with every time step and becomes computationally intractable even for small systems. Nominal

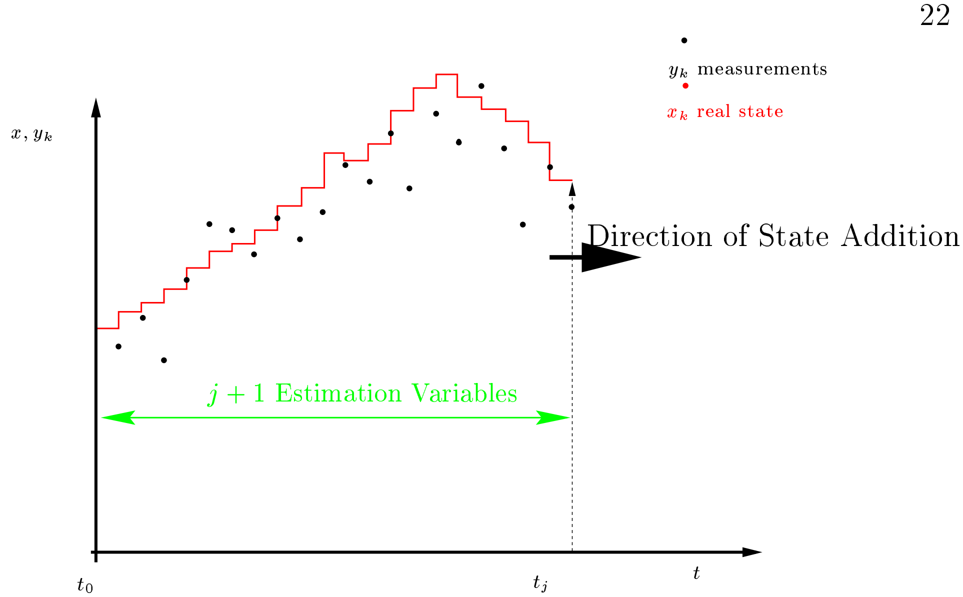

Figure 2.6: Principle of batch SE.

convergence and stability are shown in 3.2. Stability can also be deduced from the stability of the Kalman filter, due to the equivalence between the BSE and the Kalman filter (for this equivalence see 3.1).
The following theorem summarizes the stability properties of the unconstrained BSE

Theorem 2.1 (Stability of the unconstrained BSE) *The BSE given in (2.16), (2.17), with* $R^{-1} > 0$, $Q^{-1} > 0$ *and* $Q_0^{-1} > 0$ *is globally asymptotically stable and the estimate* $\hat{x}^{\star}_{k|k}$ *converges to* x_k *if* (A, C) *is detectable.*

Proof: see 3.2 and [31] □

Since the BSE gives the same result as the recursive Kalman filter, there is no advantage in using this computationally intractable formulation. However the advantage of the batch formulation lies in the possibility to incorporate prior knowledge about the system in the form of constraints which results in the constrained batch state estimator. This estimator has still the problem that the optimization problem grows with time. A solution for this are the moving or receding horizon estimators (MHE) that are discussed in 2.4.3.

Constrained Batch

The inclusion of constraints in the BSE formulation changes the structure of the estimator in the way that a change from a linear estimator to a nonlinear estimator is made.

The following state and measurements constraints are considered in addition to the equations given in (2.16), (2.17)

$$
\begin{array}{rclcrcccll}
\hat{\mathcal{X}} & = & \{\hat{x}_{j|k} \in \mathbb{R}^n & | & h_x\text{min} & \leq & H_x\hat{x}_{j|k} & \leq & h_x\text{max}, & j = 1,2,3,\dots k\} \\
\hat{\mathcal{W}} & = & \{\hat{w}_{j|k} \in \mathbb{R}^n & | & h_w\text{min} & \leq & \hat{w}_{j|k} & \leq & h_w\text{max}, & j = 0,1,2,\dots k\}
\end{array}
\tag{2.18}
$$

The fact that the first estimated state $x_{0|k}$ and the first state noise $w_{-1|k}$ are unconstrained guarantees feasibility of the connected quadratic optimization problem [31] at every time step.
If state constraints are considered for unstable A the estimator would not be able to follow the real state x_k for all possible trajectories of x_k. As a result a general nominal stability theorem for unstable A cannot be derived. This does not mean that constraints on states can not be used, but they only make sense if a guarantee that the real system does not violate these constraints can be given. One way to guarantee this would be to use a controller that forces the system to the constrained set. However the use of such a controller shifts the question about stability of the estimator to the question of the stability of the whole feedback loop. This is an even more complicated problem then the old one. Therefore no constraints on the states for unstable A are considered here.
Supplementary to (2.18), (2.18) both sets $\hat{\mathcal{X}}$, $\hat{\mathcal{W}}$ are assumed to contain the origin (for the non zero input case the resulting real state has to lay inside the constrained region). This is necessary to guarantee that the estimator converges to the value of the real state, and also to insure that the constraints form a convex region for the quadratic optimization problem.

$$
\begin{array}{rcccl}
h_x\text{min} & \leq & 0 & \leq & h_x\text{max} \\
h_w\text{min} & \leq & 0 & \leq & h_w\text{max}
\end{array}
\tag{2.19}
$$

With these definitions the following theorems about the stability of the batch state estimator can be derived

Theorem 2.2 (Nominal Asymptotic Stability of the BSE, stable A) *If A is stable then the BSE given in (2.16),(2.17) is a globally nominal asymptotically stable estimator on the constraint sets $\hat{\mathcal{X}}$ and $\hat{\mathcal{W}}$*

Proof: The proof is only outlined. For the exact derivation see Muske [31]. Feasibility of the constraints follows from the fact that no constraints are placed on the first estimates for $\hat{w}_{-1|k}$ and $\hat{x}_{0|k}$. Convergence of the estimation error sequence $e_{k|k} = x_k - \hat{x}_{k|k}$ follows from the fact that the value function for every k is bounded above by $\bar{\Psi}_k = \Psi(\{-\bar{x}_0, 0, \dots, 0\})$. This sum converges to a fixed value since A is stable and $R^{-1} > 0$. Additionally we know that our $\Psi_k^\star$ sequence is monotonically nondecreasing.

The convergence of $e_{k|k}$ follows from the stability of A and the convergence of $w^{\star}_{j|k}$ ($\Psi^{\star}_{k}$ converges and $Q^{-1} > 0$ as result $w_{j|k}$ converges to 0). Stability of $e = 0$ is guaranteed by the fact that there exists a region around the origin where the constrained BSE is equal to the unconstrained BSE. Since the unconstrained BSE is equivalent to the KF, stability of the origin is given by the stability of the KF around the origin. □

Theorem 2.3 (Nominal Asymptotic Stability of the BSE, unstable A) *If (A, C) detectable, A stable, then the BSE given in (2.16),(2.17) is a globally nominal asymptotically stable estimator on the constrained set $\hat{\mathcal{W}}$.*

Proof: The proof follows in main parts the proof given for Theorem 2.2, for details see [31]. Feasibility follows from Theorem 2.2 Convergence can be shown similar to 3.2 since the used $w_{j|k}$ sequence zeros the estimated state noise sequence for all k. Stability follows in the same way as for stable A. □

2.4.2 Kalman Filter

This linear state estimator formulation was first derived by Kalman [18]. Kalman derived this filter in his original paper using a vivid geometric approach, the orthogonal projection theory. However due to the linearity of the models used several other different and more intuitive derivation methods have been proposed. Jazwinski [16] gives a good overview on the different possibilities. He derives the Kalman filter in a straight forward way using a rigorous probabilistic setup. After that he also presents the basic ideas behind the other methods of derivations like deterministic least squares, orthogonal projection and maximum likelihood.
The Kalman filter presents a recursive solution to the batch least squares problem given in the previous section (Eq. (2.16) and (2.17)), thereby eliminating the growing computational demand during the solution of the batch problem over time. In addition to this the need to store the old measurements, like the batch method, is also removed. The new estimate is calculated using only the newest measurement and the previously calculated estimate.

The filter for time-invariant systems can be given as follows:

$$\text{Predictions:} \begin{cases} \hat{x}_{0|-1} = & \bar{x}_0 \qquad\qquad P_{0|-1} = \quad Q_0, \quad \text{Initialization} \quad k=0 \\ \hat{x}_{k|k-1} = & A\hat{x}_{k-1|k-1} \quad P_{k|k-1} = \quad AP_{k-1|k-1}A^T + Q \qquad k = 1,2,\dots \end{cases}$$

$$\begin{array}{rrcll} & \hat{x}_{k|k} & = & \hat{x}_{k|k-1} + L_k(y_k - C\hat{x}_{k|k-1}) & \\ \text{Estimates:} & P_{k|k} & = & P_{k|k-1} - L_k C P_{k|k-1} & k = 0,1,\dots \\ & L_k & = & P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1} & \end{array} \tag{2.20}$$

This estimate is the minimum variance estimate if w_j and v_j are independent, zero mean, normally distributed random variables with covariances Q and R and $\bar{x}_0$ independent normally distributed random with covariance Q_0. For linear processes it is also the most probable or maximum likelihood estimate. The covariances Q, R specify the expected magnitude of the disturbances added to the measurements and the state. With the probabilistic approach, a rigorous way to get the tuning parameters for the estimator is given if the assumption of the Gaussian nature of stochastic processes is satisfied and the covariances are known or can be quantified.

The filtered state estimate is calculated using a predicted state estimate built upon the last calculated one and the actual measurement. The matrix L_k can be seen as a linear feedback gain penalizing the difference between the actual measurement and the predicted one. L_k is calculated using the covariance of the last estimated state $P_{k|k-1}$. $P_{k|k-1}$ is calculated using a Riccati iteration with initial Condition $P_{0|-1} = Q_0$. The convergence of the Riccati iteration guarantees stability under certain assumptions.

The stability of the Kalman filter results from the iteration of the Riccati equation and is given in the following theorem:

Theorem 2.4 (Exponential Stability of the KF) *The Kalman filter is nominally exponentially stable provided (A, C) is detectable, $(A, Q^{\frac{1}{2}})$ is stabilizable, $R > 0$, $Q \geq 0$ and $Q_0 \geq 0$.*

Proof: For proofs see [7, 9] □

2.4.3 Moving Horizon Estimation

As was shown before, the batch state estimator requires the solution of a least squares problem using all of the known k previous output measurements to calculate the filtered estimate. One possible solution to reduce the amount of information processed at each step is to start with a new batch calculation at every time step by truncating the old measurement vector. An even more inferior idea would be to start a new batch estimation sequence after a specific horizon length is reached. The methods using this idea have the disadvantage that they do not employ connections between previous batches. Viable knowledge in the form of previous estimates is lost.

Jazwinski [16] proposes another idea the so called limited memory filters. These filters build up on the idea of using old covariance and mean-values to start a new batch. However, he considers only systems with measurement noise.

Another approach, the moving horizon one, tries to preserve old information by using a "information" window that slides over the measurements. The state is estimated from a horizon of the most recent $N+1$ output measurements that moves forward at each sampling time when a new measurement is available. The old information is incorporated using a startup estimate $\bar{x}_{k-N|k-N}$ that is calculated from old filtered states and a specific weight $Q_{-N|k}$. This method is discussed in a couple of papers lately. The results of these discussions are given here. The difference between the presented methods lies in the way the new initial estimate and the new initial weight are calculated, and by this how stability is guaranteed.

The moving data window reduces the problem of the growing number of decision variables in the batch state estimator to a fixed number. The first N estimates are computed using a batch estimator similar to the one presented in 2.4.1. The general idea of a moving filtering window is visualized in Figure 2.7.

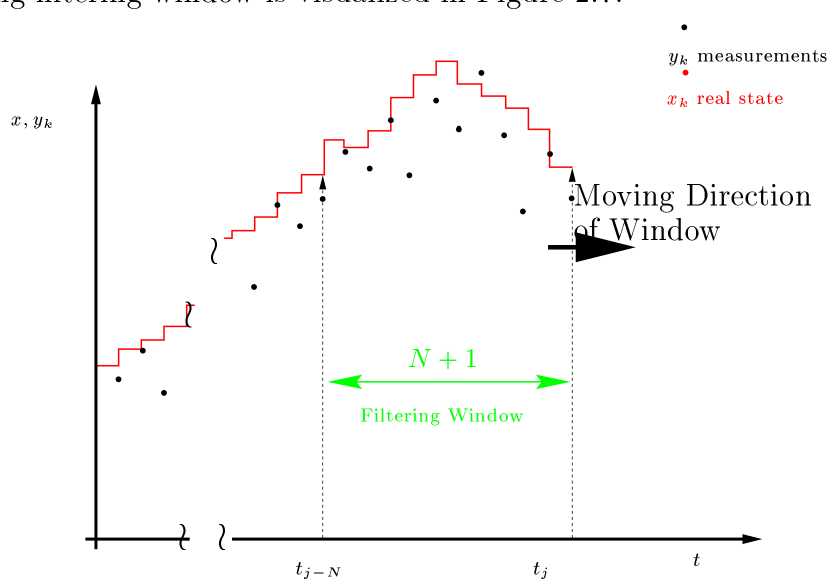

Figure 2.7: Moving filtering window, idea.

Unconstrained MHE

The first MHE formulation uses a Kalman filter like update for the initial values. This algorithm and the equivalence to the Kalman filter (similar to the equivalence of the batch estimator and the Kalman filter) was presented in an article by Muske *et al.* [37]. It also forms the the basis for the in the preceding section presented constrained MHE scheme. The startup value $\bar{x}_{k-N|k}$ is calculated from the predicted optimal estimate $\star_{k-N-1|k-N-1}$ N+1 time-steps in the past.

$$\bar{x}_{k-N|k} = A\hat{x}^{\star}_{k-N-1|k-N-1} + Bu_{k-N-1|k-N-1}$$

For the initial penalty matrix $Q_{-N|k}$ the filtering Riccati matrix $P_{k-N|k-N-1}$ resulting from the Kalman filter update as given in (2.20) with $P_{0|-1} = Q_0$ is used. Note that $P_{k-N|k-N-1}$ is the extrapolated covariance of the state estimate at time $k - N$.

$$\begin{aligned} \min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k^N &= \hat{w}^T_{k-N-1|k} P^{-1}_{k-N|k-N-1} \hat{w}_{k-N-1|k} \\ &+ \sum_{j=k-N}^{k-1} \hat{w}^T_{j|k} Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}^T_{j|k} R^{-1} \hat{v}_{j|k} \end{aligned} \tag{2.21}$$

$$\text{Subject to:} \quad \begin{aligned} \hat{x}_{k-N|k} &= \bar{x}_{k-N|k} + \hat{w}_{k-N-1|k} \\ \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + Bu_j + \hat{w}_{j|k} \\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k} \end{aligned} \tag{2.22}$$

$$\text{with:} \quad \begin{aligned} \bar{x}_{k-N|k} &= A\hat{x}^{\star}_{k-N-1|k-N-1} + Bu_{k-N-1|k-N-1} \\ P_{k-N|k-N-1} &= A(P^{-1}_{k-N-1|k-N-2} + C^T R^{-1} C)^{-1} A^T + Q \\ P_{0|-1} &= Q_0 \end{aligned}$$

Muske *et al.* [37] show that the predicted state estimate $\hat{x}^{\star}_{k+1|k}$ calculated using the scheme in (2.21) is equivalent to the Kalman filter estimate calculated by employing equation. (2.20) if the used system is autonomous:

$$\begin{aligned} x_{k+1} &= Ax_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{2.23}$$

The explanation for the equivalence is rather complicated. A more intuitive and clear derivation using the actual state estimate $x_{k|k}$ and not the predicted state $x_{k+1|k}$ is presented in section 3.1. Stability of this moving horizon scheme can be deduced from the stability of the Kalman filter using the equivalence between the Kalman filter and the MHE with Kalman filter "like" update.

Constrained Method 1, No Initial Constraints

Since the moving horizon estimator with Kalman filter update for the unconstrained case gives the same result as the pure Kalman Filter, there is no reason to use this approach due to the additional computational time that is needed to solve the optimization problem. The motivation for using the moving horizon formulation is the possibility of employing constraints. One problem that arises by doing this is the question about the set of constraints that can be used and also the type of update scheme that should be employed. This MHE formulation was given by Muske [31] and further discussed from a probabilistic viewpoint by Robertson *et al.* [41]. The idea behind this formulation is to use the stability properties of the Kalman filter in order to achieve a nominal stabilizing MHE.

This formulation uses an initial estimate $\bar{x}_{k-N|k}$ calculated by employing a Kalman filter up to time $k-N-1$ and using the Kalman predictor value $\hat{x}_{k-N|k-N-1}$ (see equation (2.20) on how to calculate it). We will denote these initial estimates with $\hat{z}_{k-N|k}$ in order to avoid confusion with old estimates. As initial weight the covariance prediction $P^{-1}_{k-N|k-N-1}$ of the Kalman filter is used.

The two previous choices guarantee the stability of the un forced estimator by employing Kalman filter stability properties. However this is not a very intuitive idea, since, as result of the used start estimates, it is in general not possible to place constraints on the initial estimates in order to keep the resulting quadratic program solvable. Therefore the initial state estimates $\hat{x}_{k-N|k-N}$ and initial state noise $\hat{w}_{k-N-1|k-N-1}$ stay unconstrained [31]. The resulting constraint sets are given in (2.25) and (2.26). Another drawback of using the Kalman filter estimates as initial guesses is that there is no connection between old MHE estimates and the newly calculated estimate, since initial values are used that could violate the constraints. Thus, this filter can be seen as an extension of the normal Kalman filter with the fact that additional constraints on the last N values are employed in order to give a "better" fitting estimate. If state constraints are considered for unstable A, the estimator would not be able to follow the real state x_k for all possible trajectories of x_k. Therefore, no constraints on the states for unstable A are utilized here. This is a restriction similar to the one that was necessary for the constrained BSE with unstable A (see 2.4.1) in order to guarantee convergence.

Additional requirements are that both sets $\hat{\mathcal{X}}$, $\hat{\mathcal{W}}$ have to contain the origin (for the non zero input case the resulting real state has to lay inside the constrained region). This is necessary to guarantee that the estimator can converge to the value of the real state and also to assure that the constraints form a convex region containing the real system state in order to keep the optimization problem quadratic. The resulting estimator with the appropriate constraints has the following structure:

$$\begin{aligned}\min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k^N &= \hat{w}_{k-N-1|k}^T P_{k-N|k-N-1}^{-1} \hat{w}_{k-N-1|k} \\ &+ \sum_{j=k-N}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \\ \hat{x}_{k-N|k} &= \hat{z}_{k-N|k} + \hat{w}_{k-N-1|k} \\ \text{Subject to:} \qquad \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + Bu_j + \hat{w}_{j|k} \\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k}\end{aligned} \tag{2.24}$$

$$\hat{\mathcal{X}} = \{\hat{x}_{j|k} \in \mathbb{R}^n \,|\, h_{x\text{min}} \leq H_x \hat{x}_{j|k} \leq h_x\text{max} \;\; j = k-N+1, k-N+2, \ldots k\} \tag{2.25}$$

$$\hat{\mathcal{W}} = \{\hat{w}_{j|k} \in \mathbb{R}^n \,|\, h_{w\text{min}} \leq \quad \hat{w}_{j|k} \leq h_w\text{max}, \; j = k-N, k-N+1, \ldots k-1\} \tag{2.26}$$

$$\text{with:} \qquad \begin{aligned}\hat{z}_{k-N|k} &= A\hat{z}_{k-N-1|k-N-1} + Bu_{k-N-1|k-N-1} \\ P_{k-N|k-N-1} &= A(P_{k-N-1|k-N-2}^{-1} + C^T R^{-1} C)^{-1} A^T + Q\end{aligned}$$

Muske [31] shows asymptotic stability for this estimator. The result is given in the following theorem.

Theorem 2.5 (Nominal Asymptotic Stability)
The estimator given by (2.24)- (2.4.3) is for stable A globally nominal asymptotic stable if (A, C) is detectable and $N \geq 1$.
For unstable A the same holds if $\hat{\mathcal{X}}$ is the whole $\mathbb{R}^{N\times n}$.

Proof: Convergence of the state estimates follows from the use of the Kalman filter estimates as updates, feasibility follows from the choice of the constraints and stability follows from the asymptotic stability of the Kalman filter around the origin. Asymptotic stability follows from the convergence, feasibility and the local stability of the estimator around the origin. For further details see [31] $\square$

Constrained Method 2, Unpenalized Initial Estimate

Another method that guarantees nominal stability was proposed by Muske and Rawlings ([35, 31]. This method allows the first estimate $\hat{x}_{k-N|k}$ to vary freely by removing the penalty Q_{-N}^{-1} on the initial estimate for the state disturbance $\hat{w}_{k-N-1|k-N-1}$. A probabilistic interpretation for this is that the value of the start estimate $\bar{x}_{k-N|k-N}$

is completely uncertain. This estimator inherits the same shortcoming as the MHE with Kalman filter start estimates. No "real" information from prior estimates or measurements from times prior $k-N$ enter the calculation. $\hat{w}_{k-N-1|k-N-1}$ can be chosen as large or small as necessary to minimize the cost function, since it does enter the cost function only indirectly. The presented realization uses a Kalman filter prediction ($\hat{z}_{k-N|k1}$) for the initial state estimate.

The resulting moving horizon optimization problem can be given as follows (as result of the free float the same constraint sets as in the previous MHE implementation can be used):

$$\min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k^N = \sum_{j=k-N}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k}$$

$$\text{Subject to:}\quad \begin{aligned} \hat{x}_{k-N|k} &= \bar{x}_{k-N|k} + \hat{w}_{k-N-1|k} \\ \bar{x}_{k-N|k} &= A\hat{x}^{\star}_{k-N-1|k-N-1} + Bu_{k-N-1|k-N-1} \\ \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + Bu_j + \hat{w}_{j|k} \\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k} \end{aligned} \tag{2.27}$$

$$\hat{\mathcal{X}} = \{\hat{x}_{j|k} \in \mathbb{R}^n \,|\, h_{x\text{min}} \leq H_x \hat{x}_{j|k} \leq h_{x\text{max}} \;\; j = k-N+1, k-N+2, \ldots k\} \tag{2.28}$$

$$\hat{\mathcal{W}} = \{\hat{w}_{j|k} \in \mathbb{R}^n \,|\, h_{w\text{min}} \leq \quad \hat{w}_{j|k} \leq h_{w\text{max}},\, j = k-N, k-N+1, \ldots k-1\} \tag{2.29}$$

$$\tag{2.30}$$

The constraints on the states are considered only if the used system model is stable, otherwise we could not guarantee that the real state stays in the constraint region, since the states can go to infinity.
However a completely new problem results from the removal of Q_{-N}. It is no longer guaranteed that a unique solution of the quadratic optimization problem exists, since we do not have a positive definite weight on $\hat{w}_{k-N-1|k}$. A unique solution can be guaranteed if the restriction to systems with (A, C) observable and horizon length bigger or equal to $n-1$ is made [31]. This is necessary in order to guarantee that the Hessian of the resulting optimization problem is positive definite. Together with the convexity of the constraints this, guarantees a unique solution.
The stability properties of the estimator are summarized in the following theorem:

Theorem 2.6 (Nominal Asymptotic Stability)
The estimator given by (2.27)- (2.29) is for stable A globally nominal asymptotic

stable on $\mathcal{X}$ and $\mathcal{W}$, if (A, C) is detectable and $N \geq n-1$.
For unstable A the same holds if no constraints on the states are enforced.

Proof: The restriction to $N \geq n-1$ guarantees the unique solution of the quadratic program. Convergence, feasibility and stability follow in a similar way as for the MHE with Kalman startup estimates. For further details see [31] □

Constrained Method 3, Prediction into the Future

MHE can be seen as the state estimation counterpart of the MPC problem, (we will come back to this in Chapter 3), therefor it seems to be logical to use similar ideas as for the ones used in MPC. For MPC stability can be guaranteed by extending the prediction horizon or control horizon infinite in the future (see for example Muske and Rawlings [33], and Scokaert and Rawlings [42]). The dual idea for MHE would be, to extend the estimation horizon in the infinite past. This is however not implementable, since only a finite number on measurement information is available. But a calculation backwards is often not possible, since this would require a invertible A. For the MPC problem this is possible, since the prediction goes into the positive time direction. The approach by Tyler and Morari [49] suggests to predict infinite in the future, instead of going in the direction of past time. Taylor and Morari consider the following system

$$\begin{aligned} x_{k+1} &= Ax_k + B_z z_k + Gw_k \\ z_{k+1} &= z_k + G_z w_{zk} \\ y_k &= Cx_k + C_z z_k + v_k \end{aligned} \tag{2.31}$$

This formulation can be reduced to the system models given earlier, however the fact that we include explicit equations describing integrated white noise in our system description will be essential for the stability of this estimator.
In the following discussion $(w_G)_{j|k}$ means the total state noise vector $[\hat{w}_{j|k}^T \; (\hat{w}_z)_{j|k}^T]^T$, a similar notation is used for the states of the expanded system x_k, z_k.
Expanding the normal GMHE cost function for the system given by (2.31) infinite in the future leads to:

$$\begin{aligned} \min_{\{(\hat{w}_G)_{k-N-1|k},\cdots,(\hat{w}_G)_{k-1|k},\cdots\}} \Psi_k^N &= (\hat{w}_G)_{k-N-1|k}^T Q_{-N}^{-1} (\hat{w}_G)_{k-N-1|k} \\ &+ \sum_{j=k-N}^{\infty} (\hat{w}_G)_{j|k}^T Q^{-1} (\hat{w}_G)_{j|k} \\ &+ \sum_{j=k-N}^{\infty} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \end{aligned} \tag{2.32}$$

To guarantee finiteness of the first infinite sum, Tyler and Morari set $(\hat{w}_G)_{j|k} = 0$ for $j \geq k$, which zeros the $\hat{w}_{k+i|k}$ prediction terms in (2.32). To guarantee convergence of the remaining part it would be possible to set $\hat{v}_{j|k} = 0$ for $j > k$. This does trivialize the problem and only makes sense in the nominal case. Instead they assume that the future measurements stay constant after k, $y_j = y_k$ for $j \geq k$. This is only possible if y_k is a steady state value y^e of (2.31). This is only possible for stable A. Nonzero steady state values y^e are then a result of the integrated noise states z_k since these stay constant if no further inputs are applied. In oder to be able to calculate the value of the sum over the v's using a Lyapunov stability result, the non-violation of the constraints on $(\hat{w}_g)_j$,$\hat{v}_j$ and $(\hat{x}_G)_j$ must be guaranteed. This is achieved by restricting the possible constraint sets. Only sets are considered for that the non violation of the constraints during $k-N, \ldots, k$ does imply that the future predictions do not violate the constraints. Therefor no state and measurement noises are considered.

Another difference between this MHE formulation and the previous one is, that the startup value for the estimator is the smoothed value $(\hat{x}_G)^\star_{k-N|k-1}$. Most other schemes use instead the the $N+1$ steps old prediction of the filtered estimate $(\hat{x}_G)^\star_{k-N|k-N-1}$. A motivation is the fact that convergence of the estimation error can be shown easier. A probabilistic motivation can not be given as easy as before. With these assumptions the MHE-S estimator can be formulated as follows.

Algorithm MHE-S

Given the old smoothed estimate $(\hat{x}_G)^\star_{k-N|k-1}$ calculate the new filtered estimate for $(\hat{x}_G)^\star_{k|k}$ as the solution of the following optimization problem:

$$\begin{aligned} \min_{\{(\hat{w}_G)_{k-N-1|k},\cdots,(\hat{w}_G)_{k-1|k},\hat{x}_{ss},\}} \Psi_k^N \;=\;& (\hat{w}_G)^T_{k-N-1|k} Q^{-1}_{-N} (\hat{w}_G)_{k-N-1|k} \\ &+ \sum_{j=k-N}^{k-1} (\hat{w}_G)^T_{j|k} Q^{-1} (\hat{w}_G)_{j|k} \\ &+ \sum_{j=k-N}^{k-1} \hat{v}^T_{j|k} R^{-1} \hat{v}_{j|k} \\ &+ ((x^e)_k - \hat{x}_{k|k})^T P ((x^e)_k - \hat{x}_{k|k}) \end{aligned} \tag{2.33}$$

with respect to the initial update:

$$\begin{aligned} \hat{x}_{k-N|k} &= \hat{x}^\star_{k-N|k-1} + \hat{w}_{k-N-1|k} \\ \hat{z}_{k-N|k} &= \hat{z}^\star_{k-N|k-1} + (\hat{w}_z)_{k-N-1|k} \end{aligned} \tag{2.34}$$

and the data update:

$$\begin{array}{rclcll} \hat{x}_{j+1|k} & = & A\hat{x}_{j|k} & + & \hat{w}_{j|k}, & j = k+1-N \ldots k \\ \hat{z}_{j+1|k} & = & \hat{z}_{j|k} & + & (\hat{w}_z)_{j|k} & \\ y_i & = & C\hat{x}_{i|k} & + & C_z\hat{z}_{i|k} + \hat{v}_{i|k} & i = k-N \ldots k \end{array} \tag{2.35}$$

subject to the following inequality constraints, (these are chosen so, that a prediction from any starting point in the constrained region without noise guarantees non violation of the constraints infinite in the future.):

$$\hat{\mathcal{X}} = \{(\hat{x}_G)_{j|k} \in \mathbb{R}^n | h_{x\text{min}} \hat{x}_{min} \leq H_x(\hat{x}_G)_{j|k} \leq h_x \text{max}, j = k-N, \ldots, k\} \tag{2.36}$$

$$\hat{\mathcal{W}} = \{(\hat{w}_G)_{j|k} \in \mathbb{R}^n | \quad h_{w\text{min}} \leq \quad H_w\hat{w}_{j|k} \leq h_w \text{max}, j = k-N-1, \ldots, k-1\} \tag{2.37}$$

$$\hat{\mathcal{V}} = \quad \{\hat{v}_{j|k} \in \mathbb{R}^n | \quad h_{v\text{min}} \leq \quad H_v\hat{v}_{j|k} \leq h_v \text{max}, j = k-N, \ldots, k\} \tag{2.38}$$

and the following equation for the steady state values:

$$x^e = x^e + B_z z_{k|k} \tag{2.39}$$

$$z^e = z_{k|k} \tag{2.40}$$

P is a solution of the following Lyapunov inequality:

$$A^T P A - P + C^T R^{-1} C \leq 0 \tag{2.41}$$

Tyler [48] gives the following theorem for the stability of the MHE-S estimation scheme

Theorem 2.7 (Nominal Asymptotic Convergence of the MHE with Infinite Prediction)
The MHE estimator stated in equations (2.33)- (2.37) for the system given by equation (2.31) converges asymptotically to the correct, real value for any horizon length $N \geq 1$ *if* $R^{-1} > 0$, $R^{-1} = R^{-T}$, $Q^{-1} > 0$, $Q^{-1} = Q^{-T}$ *and* $Q^{-1}_{-N} > 0$, $Q^{-1}_{N-1} = Q^{-T}_{N-1}$, $([A_G, [C^T C_z^T]^T)$ *is observable,* A *is stable and* P *satisfies the Lyapunov condition of eq. (2.41).*

For cases in which y_k corresponds to $x_G^\star$ values that are already violating the constraints, an additional term can be included in the infinite sum. This additional term allows to relax the constraints while still guaranteeing the convergence of the estimates. For further discussion of this idea see [48].

Remarks: It is important to notice Tyler and Morari present a method that works only for systems that have a "real" set of pure noise integrating states z_k. Furthermore these integrating states must be coupled via C_z and B_z to the measurement equation in a way, such that it can be guaranteed that a *steady state* $[x_{ss}^T\, z_{ss}^T]^T$ of the expanded system can be found while guaranteeing, that $y_k = y^e = [C^T\, C_z^T] * [x^{eT}\, z^{eT}]^T$ for all possible y_k. If this assumption is violated, there is no guarantee that the infinite sum over the predicted measurement errors $\hat{v}_{i|k} = y_k - [C^T\, C_z^T]^T * [\hat{x}_{i|k}\, \hat{z}_{i|k}]^T$ for $i \geq k$ in equation (2.32) converges. This is easy to see by setting $B_z = 0$ and $C_z = 0$. With this choice the predicted estimation error can be expressed as follows:

$$\hat{v}_{i|k} = y_k - A^i \hat{x}_{k|k} \quad i \geq k \tag{2.42}$$

Since A is stable $A^i \hat{x}_{k|k}$ converges to zero for $i \to \infty$, and thereby it follows that $\sum_{j=k-N}^{\infty} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k}$ can not converge since $y_k \neq 0$. The previous two remarks are necessary conditions for the stability of the presented estimator, however they are not explicitly stated as assumptions in the paper by Tyler and Morari. This can result in the wrong conclusion that these method works for a wider class of systems than in reality.

The presented approach can be seen as a dual to a method that allows the reduction of offsets in MPC target tracking formulations. This method was proposed recently by Muske and Rawlings [36] and reduces the influence of offsets resulting from unknown steady state inputs in a similar way to the method presented here. △

2.4.4 Luenberger Observer

Another starting point for the design of an state estimator using old estimates similar to the KF is, to consider the estimator as a dynamic system. This system is driven by the measurements via a "Controller"/filter gain. This dynamic system is designed to mimic the behavior of the real system over time. This kind of approach was first considered by Luenberger [29] and is a deterministic method. The performance criterion is to stabilize the error equation between the observer and the real system. The observer is stable if and only if all eigenvalues of the resulting error dynamics are smaller than one. The observer gain is therefore chosen so that the resulting error equation is stable.

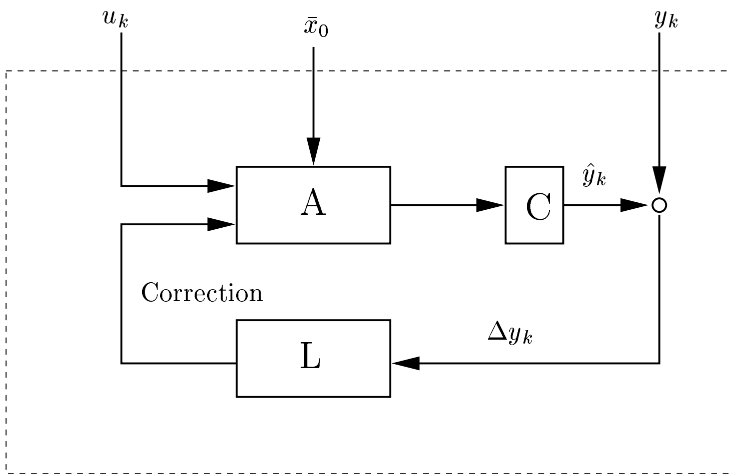

Figure 2.8: General linear estimator structure

2.4.5 Similarity of Linear Estimation Methods

For the linear unconstrained case it is important to notice that the linear estimation methods like KF, MHE, BSE and Luenberger observer can be formulated as a measurement feedback “controlled system”. The loop is closed by passing the difference between the predicted output $\hat{y}$ and the real measurements through a “linear controller”/filter gain L. The general under-laying structure has the following form:

$$\hat{x}_{k+1} = A\hat{x}_k + L(y_k - C\hat{x}_k) + Bu_k \tag{2.43}$$

The loop is closed by passing the difference between the predicted output $\hat{y}$ and the real measurements through an “linear controller”/filter gain L. The difference between these estimators is contained in L. It is an result of the different applied design methods. For the MHE with Kalman filter update, the KF and the BSE L is calculated with the goal to minimize a performance criteria. Probabilistically seen the goal is, to minimize the influence of noisy measurement and state disturbances on the system. For the Luenberger observer L is often calculated by placing the poles of the “closed loop observer”. The general under-laying structure is shown in figure 2.8

2.5 Nonlinear State Estimation

This section gives a short review about state estimation methods for the more general nonlinear systems given by equation (2.1). For several reasons, the problem of filtering for nonlinear systems is considerably more difficult, which leads to a wider variety of solution techniques than in the linear case. As opposed to linear filtering techniques, there is little theoretical justification for the use of the conditional mean as an optimal estimate, since the x_k are in general not Gaussian distributed. If, for example, the probabilistic least squares formulation derived in 2.3.2 is used to calculated the best estimate we get values that can be substantially different from the conditional mean.

Another problem (in contrast to the linear estimators) is the fact that optimal estimation methods can be derived, however they often cannot be used in a closed form, since they require the solution of infinite dimensional problems. To overcome these obstacles the application of approximation methods is necessary in order to get usable algorithms. The resulting algorithms give only suboptimal solutions. Especially if the model is strongly nonlinear these methods are only locally stable estimators, that can diverge or converge to unrealistic values over time.

First the "logical" extension of the linear Kalman filter for nonlinear case will be presented. This method is the first order extended Kalman filter that calculates the solution in a recursive manner similar to the linear formulation. Next, the extensions of the linear least squares methods for nonlinear systems are presented by stating the nonlinear batch state estimator and then the nonlinear moving horizon state estimator. These methods are motivated by the probabilistic interpretations of the estimation problem given in section 2.3.2. Finally a short overview of other nonlinear state estimation methods is given.

2.5.1 First Order Extended Kalman Filter

One straight forward method to generate an optimal nonlinear filter is to linearize the nonlinear model around a given point and to apply optimal linear state estimation to these linearized equations. A method using this idea is the first order extended Kalman filter. This estimator calculates the actual estimate by applying a linear Kalman filter on, the around the one step ahead prediction of the last filtered state of the system, linearized system equations. This method is essentially justified if the linearized system gives a good representation of the real system behavior in a sufficiently large neighborhood around the actual state of the system.

An extension to higher order filters is possible if higher order terms of the Taylor series of the system equations are considered. Here, however, only the first order extended Kalman filter is considered. For higher order extended Kalman filters and also for the exact derivation of the first order extended Kalman filter see Gelb [13] which gives an good overview about the application of nonlinear and linear filtering techniques. A derivation starting from an probabilistic viewpoint can be found in Jazwinski [16]. Further references are Stengel [47] and Lewis [28].

The extended Kalman filter considered is founded on the following discrete time systems:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k, \qquad k = 0, 1, 2, \ldots \\ y_k &= g(x_k) + v_k \end{aligned} \tag{2.44}$$

with $f,\ g \in \mathbb{C}^1$ in order to guarantee the existence of a first order Taylor series approximation.

Applying a time varying Kalman filter to the first order Taylor series yields the following algorithm for the extended Kalman filter

Prediction terms:

$$\hat{x}_{k+1|k} = f(\hat{x}_{k|k}, u_k) \tag{2.45}$$

$$\hat{P}_{k+1|k} = \mathcal{F}_k \hat{P}_{k|k} \mathcal{F}_k^T + Q \tag{2.46}$$

Observation Terms:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + L_k(y_k - g(\hat{x}_{k|k-1})) \\ L_k &= \hat{P}_{k|k-1}\mathcal{G}_k^T(\mathcal{G}_k\hat{P}_{k|k-1}\mathcal{G}_k^T + R)^{-1} \\ \hat{P}_{k|k} &= (I - L_k\mathcal{G}_k)\hat{P}_{k|k-1} \\ \mathcal{G}_k &= \frac{\partial g(x_k)}{\partial x_k}|_{x_k=\hat{x}_{k|k-1}} \\ \mathcal{F}_k &= \frac{\partial f(x_k, u_k)}{\partial x_k}|_{x_k=\hat{x}_{k|k}} \\ \hat{x}_{0|0} &= \bar{x}_0 \\ \hat{P}_{0|0} &= Q_0 \end{aligned} \tag{2.47}$$

In this formulation the new system state is predicted one step ahead by using the nonlinear system equations. With this prediction the new state estimate is calculated using a Kalman filter gain calculated using the linear approximations of g. This formulation will be later used as an update mechanism for the nonlinear MHE algorithm presented by Robertson*et al.* [41]. Yongkyu and Grizzel [45, 46]give an theorem which guarantees local stability of the EKF under some strong conditions on the system.

2.5.2 Nonlinear Batch Estimation

As was shown in section 2.3.2, the batch least squares formulation can also be applied to nonlinear systems with additive measurement and state noise for the following system structure:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k, \qquad k = 0, 1, 2, \ldots \\ y_k &= g(x_k) + v_k. \end{aligned} \tag{2.48}$$

This structure allows us to formulate a "least squares" optimization problem as the direct solution of a probabilistic formulation. The nonlinear constrained batch estimator considered here was given by Muske and Rawlings [35]. In this formulation states and system noise variables are constrained by the same set of linear constraints $\mathcal{X}$, $\mathcal{W}$ as the linear batch state estimator. These constraints are given in (2.18,2.18). With these constraints the nonlinear state estimator has the following structure similar to the linear BSE presented in 2.4.1:

$$\begin{aligned} \min_{\{\hat{w}_{-1|k},\ldots,\hat{w}_{k-1|k}\}} \Psi_k &= \hat{w}_{-1|k}^T Q_0^{-1} \hat{w}_{-1|k} + \sum_{j=0}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} \\ &+ \sum_{j=0}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \end{aligned} \tag{2.49}$$

$$\text{Subject to:} \qquad \begin{aligned} \hat{x}_{0|k} &= \bar{x}_0 + \hat{w}_{-1|k} \\ \hat{x}_{j+1|k} &= f(\hat{x}_{j|k}, u_k) + \hat{w}_{j|k} \\ y_j &= g(\hat{x}_{j|k}) + \hat{v}_{j|k} \\ \{\hat{x}_{j|k}\} \in \mathcal{X}, & \qquad \{\hat{w}_{j|k}\} \in \mathcal{W} \end{aligned} \tag{2.50}$$

The actual state estimates resulting from this minimization can be calculated by applying the calculated $\{\hat{w}^\star_{j|k}\}$ sequence to the following set of "recursive" equations:

$$\begin{aligned} \hat{x}^\star_{i+1|k} &= f(\hat{x}^\star_{i|k}, u_k) + \hat{w}^\star_{i|k} \\ \hat{x}^\star_{0|k} &= \bar{x}_0 + \hat{w}^\star_{-1|k} \end{aligned}$$

This approach has, due to the nonlinearities of the system, even more computational limitations than the linear batch state estimator. It is therefore not usable for real time applications. Muske and Rawlings give the following convergence theorem, considering initial reconstruction errors e_0 for which a feasible state noise sequence $\{w_{j|k}\}$ which zeros the reconstruction error in finite time exists:

Theorem 2.8 (Convergence of the constrained NBSE) *The by (2.49), (2.50), given BSE with $R^{-1} > 0$, $Q^{-1} > 0$, $Q_0^{-1} > 0$, $f \in \mathbb{C}^1$ converges for the nominal case to the right value x_k if the considered system is uniform observable.*

Proof: For an outlined proof see [35]. The proof bases on the same methods that are used to show the convergence of the linear batch state estimator 3.2, the only difference is that another observability condition is needed due to the fact that nonlinear systems are considered. □

2.5.3 Nonlinear Moving Horizon

Since it is not possible to calculate a closed-form recursive estimation solution for the batch estimation problem as with linear systems, it is not possible to justify a nonlinear moving horizon estimator using these results. Therefore the nonlinear moving horizon estimators (NMHE) presented here is based on the idea to keep some of the properties found earlier. These algorithms can be seen as special cases of the General Moving Horizon formulation introduced in 2.3. The difference between the presented methods here lies in the selection of the initial estimates and weights. In the linear case, the covariance of the corresponding state estimate was used to guarantee stability. Another method was to let the first estimate in the horizon to float around as needed. The methods presented here are similar to these ones.

Free Floating of Initial Estimate

This approach is a direct takeover of the linear method presented in 2.4.3. It was proposed first for nonlinear continuous time systems by Michalska and Mayne [30]. The discrete time version is taken from work by Muske and Rawlings [35], and Muske and Edgar [32].
The biggest advantage of this approach is that stability for the nominal case can be shown in a similar way as it was shown for the linear formulation, using a cost-function argument. The basic principle is that the first state disturbance estimate is not penalized in the objective function, allowing the first state estimate to be chosen freely because this state is also not included in the state and measurement constraints. One problem is that the some properties of the linear estimator with free initial estimate are also carried over. The zero weight on the initial measurement disturbance can be seen in the probabilistic framework along with the fact that we do not have any information about the initial values. This also means that no information gained in previous steps is used in order to improve the next estimate.

The NMHE estimator with unpenalized initial estimate has the following form:

$$\begin{aligned}\min_{\{\hat{w}_{k-N|k},\cdots,\hat{w}_{k|k}\}} \Psi_k^N &= \sum_{j=k-N}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} \\ &+ \sum_{j=k-N}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k}\end{aligned} \tag{2.51}$$

$$\text{Subject to:} \quad \begin{aligned}\hat{x}_{j+1|k} &= f(\hat{x}_{j|k}, u_k) + \hat{w}_{j|k} \\ y_j &= g(\hat{x}_{j|k}) + \hat{v}_{j|k} \\ \{\hat{x}_{j|k}\} \in \mathcal{X}, &\quad \{\hat{w}_{j|k}\} \in \mathcal{W}\end{aligned} \tag{2.52}$$

with the constraints defined as follows:

$$\hat{\mathcal{X}} = \{\hat{x}_{j|k} \in \mathbb{R}^n \,|\, h_{x\text{min}} \leq H_x(\hat{x}_{j|k}) \leq h_x\text{max}, j = k-N+1, k-N+2, \ldots k\} \tag{2.53}$$

$$\hat{\mathcal{W}} = \{\hat{w}_{j|k} \in \mathbb{R}^n \,|h_{w\text{min}} \leq \quad \hat{w}_{j|k} \leq h_w\text{max}, j = k-N, k-N+1, \ldots k-1\} \tag{2.54}$$

Nominal convergence of the estimation error was shown by Muske and Rawlings.

Theorem 2.9 (Convergence of the constrained NMHE with Unpenalized Initial Estimate) *The NMHE given by (2.51)- (2.54) with $R^{-1} > 0$, $Q^{-1} > 0$, $f \in \mathbb{C}^1$ converges for the nominal case to the right value x_k if the considered system is uniform observable.*

Proof: The proof follows the proof given for the linear MHE with the same update scheme. For an outline see [35]. □

Nonlinear Moving Horizon with Extended Kalman Filter Update

The in this part presented estimator employs an extended Kalman filter like update for the initial estimate $\bar{x}_{k-N|k}$. It was proposed by Robertson*et al.* [41]. Robertson *et al.*derived the update formulation using similar methods to the ones that are used for the derivation of the extended Kalman filter. The resulting algorithm calculates the initial estimates by applying to the initial weight (for the scheme discussed her $P_{k-N-1|k-N-2}$ is used for the initial estimate $Q_{-N|k}$ at time $k-1$) and initial estimates $(x_k - N \bar{-} 2|k-1)$ used one time step before, an extended Kalman filter update. The for the calculation needed linearizations are calculated by linearizing around the

smoothed estimate $\hat{x}^{\star}_{k-N-2|k-1}$. Therefor no information gained by the last estimations is lost.
The resulting estimator with the appropriate constraints has the following structure:

$$\begin{aligned}
\min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k^N &= \hat{w}^T_{k-N-1|k} P^{-1}_{k-N|k-N-1} \hat{w}_{k-N-1|k} \\
&+ \sum_{j=k-N}^{k-1} \hat{w}^T_{j|k} Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}^T_{j|k} R^{-1} \hat{v}_{j|k} \\
\hat{x}_{k-N|k} &= \bar{x}_{k-N|k} + \hat{w}_{k-N-1|k} \\
\text{Subject to:} \quad \hat{x}_{j+1|k} &= f(\hat{x}_{j|k}, u_j) + \hat{w}_{j|k} \\
y_j &= g(\hat{x}_{j|k}) + \hat{v}_{j|k}
\end{aligned} \tag{2.55}$$

with $P_{k-N|k-N-1}$ calculated using the following equations:
Observation update:

$$\begin{aligned}
\bar{x}^-_{k-N|k} &= \bar{x}_{k-N-1|k} + L_{k-N}(y_{k-N} - g(\hat{x}_{k-N-1|k-1}) - \\
&\qquad \mathcal{G}_{k-N}(\bar{x}_{k-N-1|k-1} - \hat{x}^{\star}_{k-N-1|k-1})) \\
L_{k-N} &= \hat{P}_{k-N-1|k-N-2} \mathcal{G}^T_{k-N} (\mathcal{G}_{k-N} \hat{P}_{k-N-1|k-N-2} \mathcal{G}^T_{k-N} + R)^{-1} \\
\hat{P}^-_{k-N|k-N-1} &= (I - L_{k-N} \mathcal{G}_{k-N}) \hat{P}_{k-N-1|k-N-2} \\
\mathcal{G}_{k-N} &= \frac{\partial g(x)^{\star}}{\partial x}\Big|_{x=\hat{x}^{\star}_{k-N-1|k-1}}
\end{aligned} \tag{2.56}$$

Prediction:

$$\begin{aligned}
\bar{x}_{k-N|k} &= f(\hat{x}^{\star}_{k-N-1|k-1}, u_{k-N-1}) + \mathcal{F}_{k-N}(\bar{x}^-_{k-N|k} - \hat{x}^{\star}_{k-N-1|k-1}) \\
\hat{P}_{k-N|k-N-1} &= \mathcal{F}_{k-N} \hat{P}^-_{k-N|k-N-1} \mathcal{F}^T_{k-N} + Q \\
\mathcal{F}_{k-N} &= \frac{\partial f(x,u)}{\partial x}\Big|_{x=\hat{x}^{\star}_{k-N-1|k-1}}
\end{aligned} \tag{2.57}$$

The estimation scheme is started up by using an NBSE up to time N. The additional terms in the equations are a result of use of the smoothed estimates.
Robertson *et al.* [41] show that this estimator is equivalent to the extended Kalman filter for a horizon length N of 1. The difference between this filter and the extended Kalman filter is that for horizon longer than 1 only the extended Kalman filter as initial update scheme is used. For the rest of the horizon, the nonlinear objective is minimized which results often in a better estimate. Additional constraints can be incorporated.

A disadvantage is that there exist no stability proof for this scheme. This is a result of the extended Kalman filter update, since for the extended Kalman filter stability can only be guaranteed under strong restrictions .

2.5.4 Other Nonlinear SE Methods

This part is included to give a short overview of other existing nonlinear state estimation methods. The methods presented here are not intended to be complete. They should be seen as a short overview about other methods and the basic ideas behind them.

- *Initial Estimate Approach* This method is based on the idea to use a moving horizon least squares formulation to calculate a new estimate. The difference between this and the previously presented MHE formulations is that no state errors are considered. As a result the $\hat{w}_{j|K}$ part in the GMHE formulation vanishes. What is left is a Least Squares formulation with the state estimate $\hat{x}_{k-N|k}$ as the only free parameter. Michalska and Mayne [30] show nominal convergence for this method. Implementations of this estimator can be found in articles by Bequette [2], Jang *et al.* [15] and Kim [22].

- *Observer Error Linearization* These methods are recursive algorithms using geometric methods similar to the ones used for feedback linearization. These methods are based on the idea of a local coordinate transformation such that the corresponding transformed system is linear in the reconstruction error dynamics. An estimate for the state is then calculated by using a linear observer on the linearized system and transforming the result back into the normal coordinates. An advantage of this method is that the linearization guarantees local stability of the estimate around the nominal state. The biggest drawback of this methods is that in order to calculate the transformation a set of partial differential equations must be solved. For many systems this calculation is a non-trivial task. As result this method can only be used in cases when an actual solution can be found. Figure 2.9 gives a graphical explanation of this method. This technique was independently introduced for scalar systems by Bestle and Zeitz [4] and Krener and Isidori [23]. It was extended to multivariable systems by Krener and Respondek [24] and Xia and Gao [52].

- *Extended Kalman Filter* The application of the observer error linearization involves the calculation of the necessary coordinate transformation. Since this can only be calculated easily for a very limited class of systems, other solution methods which do not require the nonlinear transformation explicitly are necessary.

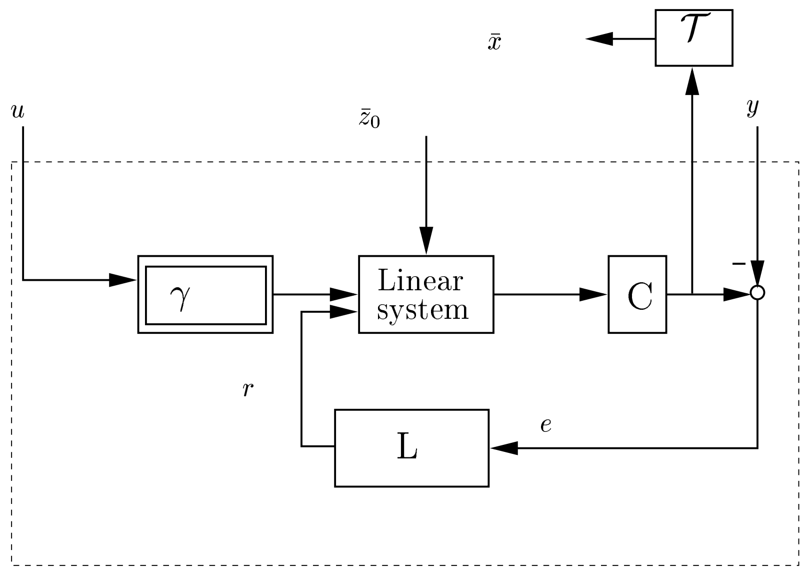

Figure 2.9: Observers basing on geometric approaches.

Birk and Zeitz [5] give a formulation that provides a solution to this problem. This formulation is referred to as a extended Luenberger observer since the observer gain is calculated by linearizing the partial differential equations. This method can be seen as the deterministic counterpart to the extended Kalman filter.

- *Sliding Mode Observer* Another nonlinear state estimation technique that gained interest in the recent time are the sliding mode observers. This method was introduced by Slotine*et al.* [44] and uses the so called "sliding surface" of the nonlinear systems to design an observer.

- *Polynomial Statistical Approximations* The extended Kalman filter utilizes a Taylor series expansion of the the system equations to calculate the next estimate. Another idea is to use polynomial approximations to characterize the system equations. These polynomial approximations allow the calculation of the mean square error in order to calculate an new estimate. Gelb [13] and Stengel [47] discuss methods based on these ideas.

Chapter 3

A Closer Look at MHE Strategies

The following chapter clarifies some of the properties of linear MHE schemes. First, a derivation concerning the connection between the Kalman filter and the batch state estimator is given. This derivation is more intuitive than the derivations by Muske and Rawlings [35] for the connection between the BSE, KF and the MHE with Kalman filter update. After this presentation, a proof showing the nominal stability of the BSE is presented. During the last part of this chapter nominal stability conditions for linear unconstrained MHE are presented.

3.1 Equivalence of MHE and BS State Estimation for Linear Systems

During the following derivations we try to clarify the connection between the generalized linear unconstrained Moving Horizon Least Squares State Estimation (GMHE), the linear Batch Least Squares State Estimation (BSE) and the recursive Kalman Filter (KF), see Figure 3.1. The work presented here is based on results derived by

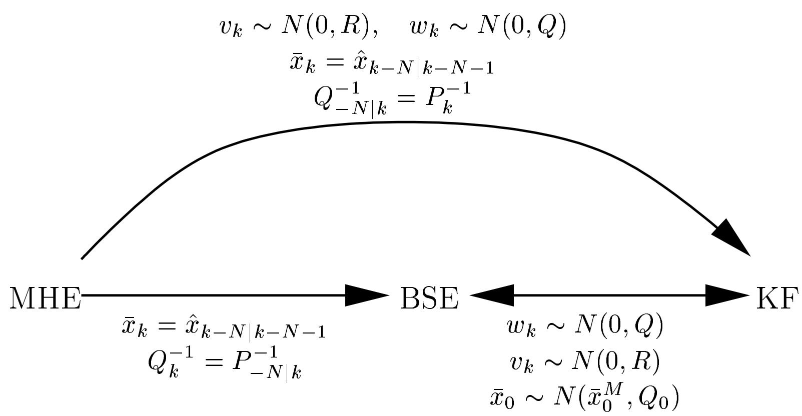

Figure 3.1: Connections between the MHE, BSE and KF.

the author and similar derivations given by Rao [38].[1]

3.1.1 The General Unconstrained Linear MHE Problem

We consider the following, time invariant, discrete time systems

$$\begin{aligned} x_{k+1} &= Ax_k + Gw_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{3.1}$$

in which $x_k \in \mathbb{R}^n$ is the state vector, $w_k \in \mathbb{R}^r$ the state noise vector, $G \in \mathbb{R}^{n\times r}$, $y_k \in \mathbb{R}^p$ the measured output and $v_k \in \mathbb{R}^p$ the measurement noise vector.
Note: the presented setup is totally deterministic. We do not assume anything about the type of disturbance entering the system via w_k or v_k. Additional assumptions will be necessary to show the connections to the KF. G offers the possibility to have the same disturbance enter different states of the system and to have a different size from x, without assuming that Q in the GMHE setup can be positive semidefinite. The generalized MHE algorithm can be stated as:

$$\min_{\{\hat{w}_{k-N-1|k},\dots,\hat{w}_{k-1|k}\}} \Psi_k : \quad \Psi_k = \hat{w}^T_{k-N-1|k} Q^{-1}_{-N|k} \hat{w}_{k-N-1|k} + \sum_{j=k-N}^{k-1} \hat{w}^T_{j|k} Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}^T_{j|k} R^{-1} \hat{v}_{j|k} \tag{3.2}$$

subject to the state equality constraints

$$\begin{aligned} \hat{x}_{k-N|k} &= \bar{x}_{k-N} + G\hat{w}_{k-N-1|k} \\ \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + G\hat{w}_{j|k} \qquad &, j = k-N \dots k-1 \\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k} \qquad &, j = k-N \dots k \end{aligned} \tag{3.3}$$

Here $R^{-1} \in \mathbb{R}^{p\times p} > 0$, $Q^{-1} \in \mathbb{R}^{r\times r} > 0$, $Q^{-1}_{-N|k} \in \mathbb{R}^{r\times r} > 0$ and $\bar{x}_{k-N}$ is an estimate of the "initial" state at time $k-N$. R^{-1} penalizes the output prediction error, Q^{-1} the estimated noise vector and $Q^{-1}_{-N|k}$ the initial estimation error. If the expected output error is small, then R^{-1} is chosen large compared to Q^{-1}. The v_j resulting from the optimization problem will be small compared to the corresponding w_j to keep the cost low. In the case of unreliable output measurements Q^{-1} is chosen large compared to R^{-1}.

During the startup phase of the GMHE ($k < N$) the BSE is used to get the first $N-1$ estimates.

[1] The author would like to express many thanks for the many helpful discussion he had with Christopher V. Rao

3.1.2 The Batch State Estimator

For N being time-variant and equal to k (3.22) leads to the Batch Least Square State Estimator.

$$\min_{\{\hat{w}_{-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k : \quad \Psi_k = \hat{w}_{-1|k}^T Q_0^{-1} \hat{w}_{-1|k} + \sum_{j=0}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \sum_{j=0}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \quad (3.4)$$

subject to the state equality constraints

$$\begin{aligned} \hat{x}_{0|k} &= \bar{x}_0 + G\hat{w}_{-1|k} \\ \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + G\hat{w}_{j|k} \quad &, j = 0 \ldots k-1 \\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k} \quad &, j = 0 \ldots k \end{aligned} \quad (3.5)$$

3.1.3 A Short Review of the Kalman Filter Equations

This section gives a short review of the discrete time Kalman Filter Equations. Suppose we have the system as given in (3.1) and the following additional assumptions are satisfied:

$$G = I; \quad w_k \sim N(0, Q); \quad v_k \sim N(0, R); \quad \bar{x}_0 \sim N(\bar{x}_0^M, Q_0) \quad (3.6)$$

Then it can be shown that x_k can be calculated as the maximum of the probability density function $p(x_k|y_0, \ldots, y_k)$ as follows:

$$\text{Predictions:} \quad \begin{cases} \hat{x}_{0|-1} = \bar{x}_0, \quad P_{0|-1} = Q_0 \quad k = 0, \quad \text{Initialization} \\ \\ \left.\begin{aligned} \hat{x}_{k|k-1} &= A\hat{x}_{k-1|k-1}, \\ P_{k|k-1} &= AP_{k-1|k-1}A^T + Q \end{aligned}\right. \quad k = 1, 2, \ldots \end{cases}$$

$$\text{Estimates:} \quad \begin{cases} \hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k(y_k - C\hat{x}_{k|k-1}) \\ P_{k|k} = P_{k|k-1} - L_k C P_{k|k-1} \qquad\qquad k = 0, 1, \ldots \\ L_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1} \end{cases}$$

For easier comparison with results in the later sections, we rewrite these equations under the usage of the following Matrix Inversion Lemma: $(B - CE^{-1}D)^{-1} = B^{-1} +$

$B^{-1}C(E-DB^{-1}C)^{-1}DB^{-1} \quad$,for B, E invertible

$$\begin{aligned}
x_{0|0} &= \bar{x}_0 + L_0(y_0 - C\bar{x}_0) \\
\hat{x}_{k|k} &= A\hat{x}_{k-1|k-1} + L_k(y_k - AC\hat{x}_{k-1|k-1}) \quad &, k = 1,2,\dots \qquad (3.7)\\
L_k &= P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1} &, k = 0,1,\dots \\
P_{0|-1} &= Q_0 \\
P_{k|k} &= (P_{k|k-1}^{-1} + CR^{-1}C^T)^{-1} & \qquad (3.8)\\
P_{k|k-1} &= A(P_{k-1|k-2}^{-1} + C^TR^{-1}C)^{-1}A^T + Q &, k = 1,2,\dots \qquad (3.9)\\
&= AP_{k-1|k-1}A^T + Q &, k = 0,1,\dots
\end{aligned}$$

3.1.4 The GMHE as a Lagrange Optimization Problem

To simplify the later derivations, it is useful to consider the GMHE problem[2] with state variables $\hat{x}_{j|k}$ and the state disturbances $\hat{w}_{j|k}$ as optimization variables.

$$\min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k},\hat{x}_{k-N|k},\cdots,\hat{x}_{k|k}\}} \Psi_k : \quad \Psi_k = \hat{w}_{k-N-1|k}^T Q_{-N|k}^{-1} \hat{w}_{k-N-1)|k} + \sum_{j=k-N}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \qquad (3.10)$$

This has to be solved subject to the equality constraints as given in (3.3). Reformulation of the equality constrained problem in Lagrange form leads to

$$\min_{Z,\tilde{\lambda}} L : L = \Psi_k - \tilde{\lambda}^T(\tilde{A}Z - b)$$

with the Lagrange multiplier $\tilde{\lambda}^T = [\ \tilde{\lambda}_0^T, \dots, \tilde{\lambda}_N^T]$, $\tilde{\lambda}_i \in \mathbb{R}^n$, where $\tilde{\lambda}_i$ is the Lagrange multiplier belonging to the equation $x_{k-N+i|k} = \dots$, with

$$Z = \begin{bmatrix} \hat{w}_{k-N-1|k} \\ \hat{x}_{k-N|k} \\ \hat{w}_{k-N|k} \\ \vdots \\ \hat{w}_{k-1|k} \\ \hat{x}_{k|k} \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} -G & I & & & & & & \\ & -A & -G & I & & & & \\ & & & -A & -G & I & & \\ & & & & \ddots & & & \\ & & & & & I & & \\ & & & & & -A & -G & I \end{bmatrix}, \quad b = \begin{bmatrix} \bar{x}_{k-N} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

[2] and by this also for the BSE problem

The necessary and sufficient conditions for a local minimizer are that the partial derivatives of L with respect to x and $\tilde{\lambda}$ vanish.

$$\begin{aligned} \frac{\partial L}{\partial Z} &= 0 \\ \frac{\partial L}{\partial \tilde{\lambda}} &= 0 \end{aligned}$$

This leads to

$$\begin{bmatrix} \widetilde{QC} & -\tilde{A}^T \\ \tilde{A} & 0 \end{bmatrix} \begin{bmatrix} Z \\ \lambda \end{bmatrix} = \begin{bmatrix} \tilde{Y} \\ b \end{bmatrix}, \qquad \lambda = \frac{\tilde{\lambda}}{2}$$

with

$$\widetilde{QC} = \begin{bmatrix} Q^{-1}_{-N|k} & & & & & & \\ & C^T R^{-1} C & & & & & \\ & & Q^{-1} & & & & \\ & & & C^T R^{-1} C & & & \\ & & & & \ddots & & \\ & & & & & Q^{-1} & \\ & & & & & & C^T R^{-1} C \end{bmatrix}, \quad \tilde{Y} = \begin{bmatrix} 0 \\ C^T R^{-1} y_{k-N} \\ 0 \\ C^T R^{-1} y_{k-N+1} \\ \vdots \\ 0 \\ C^T R^{-1} y_k \end{bmatrix}$$

After reordering of the equations and unknowns with

$$\tilde{Z}^T = \begin{bmatrix} \hat{w}^T_{k-N-1|k} & \hat{x}^T_{k-N|k} & \lambda^T_0 & \hat{w}^T_{k-N|k} & \hat{x}^T_{k-N+1|k} & \lambda^T_1 & \cdots & \hat{w}^T_{k-1|k} & \hat{x}^T_{k|k} & \lambda^T_k \end{bmatrix}$$

we end up with the following linear algebra problem for the solution of (3.10):

$$
\begin{bmatrix}
Q^{-1}_{-N|k} & 0 & G^T & & & & & & & & & & \\
0 & C^TR^{-1}C & -I & 0 & 0 & A^T & & & & & & & \\
-G & I & 0 & 0 & 0 & 0 & & & & & & & \\
 & 0 & 0 & Q^{-1} & 0 & G^T & & & & & & & \\
 & 0 & 0 & 0 & C^TR^{-1}C & -I & 0 & 0 & A^T & & & & \\
 & -A & 0 & -G & I & 0 & 0 & 0 & 0 & & & & \\
 & & & & 0 & 0 & Q^{-1} & 0 & G^T & & & & \\
 & & & & 0 & 0 & 0 & C^TR^{-1}C & -I & 0 & 0 & A^T & \\
 & & & & -A & 0 & -G & & 0 & 0 & 0 & 0 & \\
 & & & & & & & \ddots & & & & & \\
 & & & & & & & Q^{-1} & 0 & G^T & & & \\
 & & & & & & & 0 & C^TR^{-1}C & -I & 0 & 0 & A^T \\
 & & & & & -A & 0 & -G & I & 0 & 0 & 0 & 0 \\
 & & & & & & & & 0 & 0 & Q^{-1} & 0 & G^T \\
 & & & & & & & & 0 & 0 & 0 & C^TR^{-1}C & -I \\
 & & & & & & & & -A & 0 & -G & I & 0
\end{bmatrix}
\begin{bmatrix}
\hat{w}_{k-N-1|k} \\ \hat{x}_{k-N|k} \\ \lambda_0 \\ \hat{w}_{k-N|k} \\ \hat{x}_{k-N+1|k} \\ \lambda_1 \\ \hat{w}_{k-N+1|k} \\ \hat{x}_{k-N+2|k} \\ \lambda_2 \\ \vdots \\ \hat{w}_{k-2|k} \\ \hat{x}_{k-1|k} \\ \lambda_{N-1} \\ \hat{w}_{k-1|k} \\ \hat{x}_{k|k} \\ \lambda_N
\end{bmatrix}
=
\begin{bmatrix}
0 \\ C^TR^{-1}y_{k-N} \\ \bar{x}_{k-N} \\ 0 \\ C^TR^{-1}y_{k-N+1} \\ 0 \\ 0 \\ C^TR^{-1}y_{k-N+2} \\ 0 \\ \vdots \\ 0 \\ C^TR^{-1}y_{k-1} \\ 0 \\ 0 \\ C^TR^{-1}y_k \\ 0
\end{bmatrix}
\tag{3.11}
$$

The banded structure of this problem is an indicator that we are able to solve it using a recursive algorithm. Using the fact that the states are only loosely coupled we derive the connection between the MHSE, the BSE and the Kalman Filter (for the special case that we assume that w_k and v_k are random variables with zero mean and that $G = I$).

3.1.5 Connections between the BSE and the KF

During the following part the connections between the BSE and the KF are outlined. The solution of the BSE problem can be calculated from equation (3.11) under the additional assumption that $N = k$.

Special case $k = 0$, Startup

Using the results in section 3.1.4 derived equation (3.11) with $N = k = 0$ we have to solve the following problem to get the optimal estimate $\hat{x}_{0|0}$ for the BSE

$$\begin{bmatrix} Q_0^{-1} & 0 & G^T \\ 0 & C^T R^{-1} C & -I \\ -G & I & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_{-1|0} \\ \hat{x}_{0|0} \\ \lambda_0 \end{bmatrix} = \begin{bmatrix} 0 \\ C^T R^{-1} y_0 \\ \bar{x}_0 \end{bmatrix} \tag{3.12}$$

After block elimination and the additional assumption that $I + C^T R^{-1} C G Q_0 G^T$ is invertible the following can be derived

$$\begin{bmatrix} Q_0^{-1} & 0 & G^T \\ 0 & 0 & -(I + C^T R^{-1} C G Q_0 G^T) \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_{-1|0} \\ \hat{x}_{0|0} \\ \lambda_0 \end{bmatrix} =$$
$$\begin{bmatrix} 0 \\ C^T R^{-1} y_0 - C^T R^{-1} C \bar{x}_0 \\ \bar{x}_0 + G Q_0 G^T (I + C^T R^{-1} C G Q_0 G^T)^{-1} C^T R^{-1} (y_0 - C \bar{x}_0) \end{bmatrix} \tag{3.13}$$

This leads for $\hat{x}_{0|0}$ to

$$\hat{x}_{0|0} = \bar{x}_0 + G Q_0 G^T (I + C^T R^{-1} C G Q_0 G^T)^{-1} C^T R^{-1} (y_0 - C \bar{x}_0)$$

Assuming $G Q_0 G^T$ is invertible we can simplify the equation to

$$\hat{x}_{0|0} = \bar{x}_0 + ((G Q_0 G^T)^{-1} + C^T R^{-1} C)^{-1} C^T R^{-1} (y_0 - C \bar{x}_0)$$

After application of the matrix inversion lemma

$$B^{-1} C (E + D B^{-1} C) = (B + C E^{-1} D) C E^{-1} \qquad , B, E \quad \text{invertible}$$

the following for $\hat{x}_{0|0}$ can be established

$$\hat{x}_{0|0} = \bar{x}_0 + G Q_0 G^T C^T (R + C G Q_0 G^T C^T)^{-1} (y_0 - C \bar{x}_0)$$

This equals the Kalman Filter equation for $\hat{x}_{0|0}$, if $G = I^{n \times n}$.

Connection between BSE and Kalman Filter, $k >= 1$

For k=1 block elimination of the resulting system yields

$$\begin{bmatrix} Q_0^{-1} & 0 & G^T & & & \\ & 0 & -GQ_0G^T\Big((GQ_0G^T)^{-1}+C^TR^{-1}C\Big) & 0 & 0 & A^T \\ & I & 0 & 0 & 0 & \Big((GQ_0G^T)^{-1}+C^TR^{-1}C\Big)^{-1}A^T \\ & & & Q^{-1} & 0 & G^T \\ & & & 0 & C^TR^{-1}C & -I \\ & & & -G & I & A\Big((GQ_0G^T)^{-1}+C^TR^{-1}C\Big)^{-1}A^T \end{bmatrix} \begin{bmatrix} \hat{w}_{-1|1} \\ \hat{x}_{0|1} \\ \lambda_0 \\ \hat{w}_{0|1} \\ \hat{x}_{1|1} \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 0 \\ C^TR^{-1}y_0 - C^TR^{-1}C\bar{x}_0 \\ \hat{x}_{0|0} \\ 0 \\ C^TR^{-1}y_1 \\ A\hat{x}_{0|0} \end{bmatrix} \tag{3.14}$$

In this form we have achieved independence of the unknowns of $\hat{x}_{1|1}$ (right lower corner) from the unknowns of $\hat{x}_{0|1}$ (left upper corner).
To stay consistent with the KF we define

$$\tilde{P}_{0|-1}^{-1} = Q_0^{-1}$$

and

$$\tilde{P}_{0|0}^{-1} = \Big((G\tilde{P}_{0|-1}^{-1}G^T)^{-1} + C^TR^{-1}C\Big)^{-1}$$

see equations (3.8) and (3.9).

$$\begin{bmatrix} \tilde{P}_{0|-1}^{-1} & 0 & G^T & & & \\ & 0 & -G\tilde{P}_{0|-1}^{-1}G^T\tilde{P}_{0|0} & 0 & 0 & A^T \\ & I & 0 & 0 & 0 & \tilde{P}_{0|0}^{-1}A^T \\ & & & Q^{-1} & 0 & G^T \\ & & & 0 & C^TR^{-1}C & -I \\ & & & -G & I & A\tilde{P}_{0|0}^{-1}A^T \end{bmatrix} \begin{bmatrix} \hat{w}_{-1|1} \\ \hat{x}_{0|1} \\ \lambda_0 \\ \hat{w}_{0|1} \\ \hat{x}_{1|1} \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 0 \\ C^TR^{-1}y_0 - C^TR^{-1}C\bar{x}_0 \\ \hat{x}_{0|0} \\ 0 \\ C^TR^{-1}y_1 \\ \hat{x}_{1|0} \end{bmatrix} \tag{3.15}$$

Remarks: Note that $\tilde{P}$ and P coincide for $G = I$. △

Applying now the following variable transformation

$$-G\hat{\tilde{w}}_{0|1} = -G\hat{w}_{0|1} + A\tilde{P}_{0|0}^{-1}A^T\lambda_1$$

and using only the lower right block of the resulting equations we get

$$\begin{bmatrix} P_{1|0}^{-1} & 0 & G^T \\ 0 & C^TR^{-1}C & -I \\ -G & I & 0 \end{bmatrix} \begin{bmatrix} \hat{\tilde{w}}_{0|1} \\ \hat{x}_{1|1} \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 0 \\ C^TR^{-1}y_1 \\ \hat{x}_{1|0} \end{bmatrix} \tag{3.16}$$

with

$$P_{1|0}^{-1} = G^T(A\tilde{P}_{0|0}^{-1}A^T + GQG^T)^{-1}G$$

These equations have the same structure as the set of linear equations we got for $k = 0$. The major difference is that Q_0^{-1} is replaced by $\tilde{P}_{1|0}^{-1}$ and $\bar{x}_0$ by $\hat{x}_{1|0}$. $\tilde{P}_{1|0}^{-1}$ is calculated using $\tilde{P}_{0|0}$. Block elimination as in section 3.1.5 results in the following equation for $\hat{x}_{1|1}$

$$\hat{x}_{1|1} = \hat{x}_{1|0} + G\tilde{P}_{1|0}G^TC^T(R + CG\tilde{P}_{1|0}G^TC^T)^{-1}(y_1 - C\hat{x}_{1|0})$$

It is clear from the above derivation, that for every $k > 1$, we can successively repeat the above outlined steps with the transformation

$$-G\hat{\tilde{w}}_{j|k} = -G\hat{w}_{j|k} + A\tilde{P}_{j|j}^{-1}A^T\lambda_1$$

For example for k=2 we would end up with the following system:

$$\begin{bmatrix} \tilde{P}_{0|-1}^{-1} & 0 & G^T & & & & & & \\ & 0 & -G\tilde{P}_{0|-1}^{-1}G^T\tilde{P}_{0|0} & 0 & 0 & A^T & & & \\ & I & 0 & 0 & 0 & \tilde{P}_{0|0}^{-1}A^T & & & \\ & & & \tilde{P}_{1|0}^{-1} & 0 & G^T & & & \\ & & & & 0 & -G\tilde{P}_{1|0}^{-1}G^T\tilde{P}_{1|1} & 0 & 0 & A^T \\ & & & & I & 0 & 0 & 0 & \tilde{P}_{1|1}^{-1}A^T \\ & & & & & & \tilde{P}_{2|1}^{-1} & 0 & G^T \\ & & & & & & 0 & C^TR^{-1}C & -I \\ & & & & & & -G & I & A\tilde{P}_{1|1}^{-1}A^T \end{bmatrix} \begin{bmatrix} \hat{w}_{-1|2} \\ \hat{x}_{0|2} \\ \lambda_0 \\ \hat{w}_{0|2} \\ \hat{x}_{1|2} \\ \lambda_1\hat{w}_{1|2} \\ \hat{x}_{2|2} \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ C^TR^{-1}y_0 - C^TR^{-1}C\bar{x}_0 \\ \hat{x}_{0|0} \\ 0 \\ C^TR^{-1}y_1 - C^TR^{-1}C\hat{x}_{1|0} \\ \hat{x}_{1|1} \\ 0 \\ C^TR^{-1}y_2 \\ \hat{x}_{2|1} \end{bmatrix} \tag{3.17}$$

Since this is possible for every k, it follows, that

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + G\tilde{P}_{k|k-1}G^TC^T(R + CG\tilde{P}_{k|k-1}G^TC^T)^{-1}(y_k - C\hat{x}_{k|k-1})$$

This is equivalent to the solution of the Kalman Filter for $G = I$ and v_k and w_k are independent zero mean random variables.

3.1.6 Connection between BSE and the GMHE

To derive the relation between the BSE and the MHSE we look at equation 3.11 for $k > N$. We see, that if we choose Q_k^{-1} as

$$Q_k^{-1} = P_k^{-1} = G^T(A((GP_{k-1}G^T)^{-1} + C^TR^{-1}C)^{-1}A^T + GQG^T)^{-1}G \tag{3.18}$$

and $\bar{x}_{k-N}$ as

$$\bar{x}_{k-N} = \hat{x}_{k-N|k-N-1} \tag{3.19}$$

Then we end up with the Batch Estimator equations starting at the k's block. With this we have established that the Batch estimator is equivalent to the Moving Horizon Estimator for the specific choice of Q_k^{-1} and $\bar{x}_{k-N}$.

3.2 Stability of the Linear Batch Estimator

As we have seen in the previous part there is a close connection between the Kalman filter, BSE and the MHE. The following part presents a nominal stability proof for the BSE and also gives stability conditions for the Kalman filter and the MHE with Kalman update. This proofs are based on results presented by Rawlings [39], Muske and Rawlings [37] and Souza *et al.* [9].

Definition 3.1 (Nominal Asymptotic Stability) *We say that an estimator is nominal asymptotic stable if the reconstruction error of the estimate converges to zero for an arbitrary non-zero initial condition without state or measurement disturbances.*

Definition 3.2 (Reconstruction error) *The reconstruction error e_k is defined as the difference between the actual state and the estimated state,*

$$e_k = x_k - \hat{x}_{k|k} \tag{3.20}$$

We consider the state estimation problem for the following linear discrete time system

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \qquad k = 0, 1, 2, ... \\ y_k &= Cx_k + v_k \end{aligned} \tag{3.21}$$

with $x_k \in \mathbb{R}^n$,$w_k \in \mathbb{R}^r$, $y_k \in \mathbb{R}^p$, $v_k \in \mathbb{R}^p$. Since we are interested in the nominal stability we only consider autonomous systems without disturbances, which means

$w_k = v_k = 0$.
We consider the following deterministic batch state estimator

$$\min_{\{\hat{w}_{-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k : \quad \Psi_k = \hat{w}^T_{-1|k} Q^{-1}_{-N|k} \hat{w}_{-1|k} + \sum_{j=0}^{k-1} \hat{w}^T_{j|k} Q^{-1} \hat{w}_{j|k} + \sum_{j=0}^{k} \hat{v}^T_{j|k} R^{-1} \hat{v}_{j|k} \quad (3.22)$$

subject to the state equality constraints

$$\begin{aligned} \hat{x}_{0|k} &= \bar{x}_0 + \hat{w}_{-1|k} \\ \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + \hat{w}_{j|k} \quad , j = k-N \dots k-1 \\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k} \quad , j = k-N \dots k \end{aligned} \quad (3.23)$$

This method is basically an optimal least squares fit of the system trajectory with respect to the given measurements.
We assume that the following holds:

A6 $Q^{-1} > 0$, $R^{-1} > 0$, $Q_0^{-1} > 0$

The following theorem summarizes the stability properties of the BSE for the given system.

Theorem 3.1 Asymptotic stability of the BSE *If assumption A6 holds and additionally (A, C) detectable, then e_k converges asymptotically to 0 and the by (3.22) given state estimator with $G = I$ for (3.21) is nominal stable.*

Proof: The proof can be divided in 4 parts

1. show that $\Psi^\star_k$ is monotonically nondecreasing
2. existence of upper bound for $\Psi^\star_k$
3. part 1 and part 2 imply that $\lim_{k\to\infty} e^\star_k \to 0$
4. stability of the reconstruction error around 0

1. $\Psi^\star_k$ is monotonically nondecreasing

Let $\Psi^\star_{k+1}$ be optimal at time k+1 with the associated state error sequence $\{\hat{w}_{j|k+1}\}$

$$\Psi^\star_{k+1}(\{\hat{w}_{-1|k+1}, \hat{w}^\star_{0|k+1}, \dots, \hat{w}^\star_{k|k+1}\}) \geq 0$$

If we now use this $w_{j|k+1}$ sequence to calculate $\tilde{\Psi}_k$ we know from the optimality theorem that this value is non-optimal

$$\tilde{\Psi}_{k+1} = \Psi^\star_{k+1} - \hat{w}^{\star T}_{k|k+1} Q^{-1} \hat{w}^\star_{k|k+1} - \hat{v}^{\star T}_{k+1|k+1} R^{-1} \hat{v}^\star_{k+1|k+1} \geq \Psi^\star_k$$

which shows that

$$\Psi^\star_{k+1} \geq \Psi^\star_k \quad \Leftrightarrow \quad \{\Psi^\star_j\} \quad \text{is monotonically nondecreasing}$$

2. upper bound for $\Psi^\star_k$

Assuming we allow only $\hat{w}_{-1|j}$ to have an value different from 0, which implies that $\hat{w}_{i|j} = 0 \; \forall i \in \{0, 1, \dots j-1\}$, and $\hat{v}_{i|j} = 0 \; \forall i \in \{0, 1, \dots j\}$.
We get as the only possible solution to our optimization problem $\hat{w}_{-1|k} = x_0 - \bar{x}_0$ with

$$\Psi_{max} = (x_0 - \bar{x}_0)^T Q_0^{-1} (x_0 - \bar{x}_0) > \Psi^\star_j \geq 0$$

This solution guarantees an upper bound, because by the principle of optimality for linear systems, less degrees of freedom are associated with an higher cost. Furthermore it is importend to notice that with this choice the reconstruction error equation $e_{j|k} = Ae_{j-1|k} - \hat{w}_{j-1|k} = 0 \; \vee j = 1, ...k$ is equal to 0 for all times.

3. $\lim_{k\to\infty} e^\star_k \to 0$

From 1. we know

$$\Psi^\star_{k+1} - \Psi^\star_k \geq \hat{w}^{\star T}_{k|k+1} Q^{-1} \hat{w}^\star_{k|k+1} - \hat{v}^{\star T}_{k+1|k+1} R^{-1} \hat{v}^\star_{k+1|k+1} \geq 0$$

using the existence of an upper bound (2.) we get

$$\lim_{k\to\infty} \Psi^\star_{k+1} - \Psi^\star_k \to 0 \quad \overset{A6}{\Longleftrightarrow} \quad \lim_{k\to\infty} \hat{w}^{\star T}_{k|k+1} Q^{-1} \hat{w}^\star_{k|k+1} + \hat{v}^{\star T}_{k+1|k+1} R^{-1} \hat{v}^\star_{k+1|k+1} \to 0$$

Since $Q^{-1} > 0$ and R^{-1} we can separate the 2 parts of the last sum and get the following two for the convergence of the sum sufficient conditions

$$\lim_{k\to\infty} \hat{w}^{\star T}_{k|k+1} Q^{-1} \hat{w}^\star_{k|k+1} \to 0$$
$$\lim_{k\to\infty} \hat{v}^{\star T}_{k+1|k+1} R^{-1} \hat{v}^\star_{k+1|k+1} \to 0$$

which are equivalent to $\lim_{k\to\infty} \hat{w}^\star_{k|k+1} \to 0$ and $\lim_{k\to\infty} \hat{v}^\star_{k+1|k+1} \to 0$.
Using this conditions for large k on the system equations (assuming that $\hat{w}^\star_{k|k+1}$ and $\hat{v}^\star_{k+1|k+1}$ have already small values) we get:

$$\begin{aligned} \hat{x}^\star_{k-N|k} &= A\hat{x}^\star_{k-1|k} \\ y_k &= C\hat{x}^\star_{k|k} \end{aligned} \tag{3.24}$$

These equations have the same form as our system without disturbance (3.1.1). We also know that y_k is the same for both equations. To conclude from the same y_k that we have the same x_k for an unforced system we need the additional assumption that (A, C) must be detectable. This assumption guarantees that the unobservable modes approaches zero for the unforced system and that the observable modes are the same after an long enough time.

4. Stability of the reconstruction error $e = 0$

The stability of the reconstruction error point $e = 0$ follows from the fact that the Kalman filter is stable around the origin. With this follows, using the equivalence of the Kalman filer and the batch state estimator, that the reconstruction error for the batch estimator is stable around the origin. This concludes together with the convergence of the value function our proof since convergence and stability imply asymptotic stability. □

3.3 Nominal Stability of the Linear Unconstrained MHE with Prediction Update

In the following section nominal stability conditions for the unconstrained MHE with prediction update for $\bar{x}_{k-N|k}$ and for any $Q_{-N|k} = Q_{-N} = Q_{-N}^T = \text{const} > 0$ are derived. Prediction update means that a one-step ahead prediction of an from past estimates derived $\bar{x}_{k-N-1}$ is used to calculate the initial guess for $\bar{x}_{k-N|k}$:

$$\bar{x}_{k-N|k} = A\bar{x}_{k-N-1}$$

The usage of this prediction is motivated by the MHE with KF like update given in section 2.4.3. However in contrary to this the here presented approach uses an fixed positive definite and symmetric initial weight Q_{-N}. The major result of this section will be sufficient tests which allow to check if a specific Q_{-N} leads to a nominal stable state estimator.

Possible values for $\bar{x}_{k-N-1}$, are for example, the last smoothed estimate $\hat{x}^{\star}_{k-N-1|k-1}$ used in the MHE-S approach (2.33) or the filtered estimate $\hat{x}^{\star}_{k-N-1|k-N-1}$ used in the MHE formulation given in section 2.4.3. A discussion of advantages and disadvantages of the different updates will be give in section 3.3.3.

First the similarity between the MHE formulation and the linear unconstrained finite horizon model predictive control (MPC) problem is shown. This seems appropriate since many authors state the equivalence between the MHE and MPC methods without further discussion. It will be shown that the MPC and MHE formulation for nominal initial errors partly, but not totally coincide. In order to show this equivalence it is necessary to rewrite the MHE formulation in terms of the estimation error $e_{j|k} = x_j - \hat{x}_{j|k}$. Use of this error equation transforms the MHE equations in a comparable form to the MPC formulation.

Motivated by the dynamic programming solution for the MPC/LQR problem, a solution of the unconstrained MHE is calculated. The usage of dynamic programming allows further insight in the connection between MPC and MHE. A major difference between these two schemes is that in the MPC formulation the optimization must be solved only "backwards" until $u_{k|k}$ is known. The MHE problem in contrary requires additionally to iterate "forward" once $\hat{x}^{\star}_{k-N|k}$ is known. This is necessary to derive the equations for the new filtered estimate $\hat{x}^{\star}_{k|k}$. The solution generated by solving this optimization problem allows to derive a nominal stability test for the MHE.

3.3.1 Equivalence between MHE and MPC

Most MHE formulations approaches can been motivated as the state estimation counterpart of linear MPC [41, 37, 32]. Model predictive control is a control strategy, in which the controller determines an open loop input profile $\{u_{k|k}, \dots u_{k+N-1|k}\}$. This profile optimizes a given performance objective over a time interval spanning from the current time to the prediction horizon length N in the future. The first element of the resulting input sequence is implemented and the whole process is repeated for the next control step. A typical formulation for the unconstrained finite horizon MPC or LQ problem is [39]:

$$\min_{\{u_{k|k},\dots,u_{k+N-1|k}\}} \Phi_k^N : \quad \Phi_k^N = \sum_{j=0}^{N-1} \left(x^T_{k+j|k}\tilde{Q}x_{k+j|k} + u^T_{k+j|k}\tilde{R}u_{k+j|k} \right) + x^T_{k+N|k}\tilde{Q}_N x_{k+N|k} \tag{3.25}$$

With $\tilde{Q} = \tilde{Q}^T \geq 0 \in \mathbb{R}^{n\times n}$, $\tilde{Q}_N^T = \tilde{Q}_N \geq 0 \in \mathbb{R}^{n\times n}$ and $\tilde{R} = \tilde{R}^T > 0 \in \mathbb{R}^{m\times m}$ subject to the state equality constraints

$$\begin{aligned} x_{j+1|k} &= Ax_{j|k} + u_{j|k} \quad ,j = k,\dots,N-1 \\ x_{k|k} &= x_k \\ u_k &= u_{k|k} \end{aligned} \tag{3.26}$$

Since no additional constraints on states or inputs are enforced, the MPC problem coincides with the traditional LQ problem [17]. It can be shown [31, 7] that for $N \to \infty$ (classical LQR) the following holds

Theorem 3.2 (Stability infinite LQ problem) *The controller given by equations (3.25)-(3.26) with $\tilde{Q} \geq 0$, $\tilde{R} > 0$, $N \to \infty$ is exponentially stabilizing, if $[A, \tilde{Q}^{1/2}]$ is observable and $[A, B]$ is stabilizable. Furthermore the optimal value of $\Phi = \Phi^\star(x_k)$ can be calculated using the following formulation:*

$$\Phi^\star = x_k^T \Pi_\infty x_k$$

with Π_∞ as the solution of the following DARE (discrete algebraic Riccati equation):

$$\Pi_\infty = \tilde{Q} + A^T \Pi_\infty A - A^T \Pi_\infty B(\tilde{R} + B^T \Pi_\infty B)^{-1} B^T \Pi_\infty A \tag{3.27}$$

The resulting steady state gain matrix K_∞ is given by

$$\begin{aligned} K_\infty &= -(\tilde{R} + B^T \Pi_\infty B)^{-1} B^T \Pi_\infty A \\ u_k &= K_\infty x_{x|k} = K_\infty x_k \end{aligned}$$

Proof: See [31, 7]. □

Now the MHE given by equations (3.2),(3.3) with prediction update for $\bar{x}_{k-N|k}$ is considered. The general formulation for $\bar{x}_{k-N-1}$ is kept for the moment, specific values and therefrom resulting properties will be considered later.

$$\begin{aligned} \min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k^N : \quad \Psi_k^N =& \hat{w}_{k-N-1|k}^T Q_{-N}^{-1} \hat{w}_{k-N-1|k} \\ &+ \sum_{j=0}^{N-1} \left(\hat{w}_{k-N+j|k}^T Q^{-1} \hat{w}_{k-N+j|k} + \hat{v}_{k-N+j|k}^T R^{-1} \hat{v}_{k-N+j|k} \right) \\ &+ \hat{v}_{k|k}^T R^{-1} \hat{v}_{k|k} \end{aligned} \tag{3.28}$$

subject to

$$\begin{aligned} \hat{x}_{k-N|k} &= A\bar{x}_{k-N-1} + \hat{w}_{k-N-1|k} \\ \hat{x}_{j+1|k} &= A\hat{x}_{j|k} + \hat{w}_{j|k} & , j = k-N \dots k-1 \\ y_j &= C\hat{x}_{j|k} + \hat{v}_{j|k} & , j = k-N \dots k \end{aligned} \tag{3.29}$$

If the equations (3.28),(3.29) for the MHE and the equations (3.25)-(3.26) for the MPC are compared, they seem to be different. The major difference between these two schemes is that in the MPC formulation an optimization over a cost function containing the inputs u and the states x with respect to the inputs u must be solved. In the MHE formulation the state disturbances $\hat{w}$ replace the inputs u. Additionally there are three other variables in the cost function, the state estimates $\hat{x}$, the output disturbances $\hat{v}$ and the output measurements y. Since the goal here is to guarantee nominal stability[3] of the MHE the $y's$ can be eliminated using the nominal system equations:

$$y_j = Cx_j$$

The resulting estimation problem can be seen as a deterministic recovery problem from a wrong initial state. $\hat{w}_{j|k}$ and $\hat{v}_{j|k}$ take now the place of "curve fitting parameters." If non-nominal system behavior is considered, the MHE formulation can be seen as the counterpart to robust MPC formulations; therefore an easy comparison of MHE and MPC is only possible for the nominal case.
In order to clarify the connections, we start up by reformulating (3.28)-(3.29) in terms of the estimation error $e_{j|k}$,

$$e_{j|k} = x_j - \hat{x}_{j|k}, \; e_k = x_k - \hat{x}_{k|k} \tag{3.30}$$

so that

$$\begin{aligned} \min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k^N : \quad \Psi_k^N =& \hat{w}_{k-N-1|k}^T Q_{-N}^{-1} \hat{w}_{k-N-1|k} \\ &+ \sum_{j=0}^{N-1} \left(e_{k-N+j|k}^T C^T R^{-1} C e_{k-N+j|k} + \hat{w}_{k-N+j|k}^T Q^{-1} \hat{w}_{k-N+j|k} \right) \\ &\quad + e_{k|k}^T C^T R^{-1} C e_{k|k} \end{aligned} \tag{3.31}$$

subject to

$$e_{k-N|k} = A\bar{e}_{k-N-1} - \hat{w}_{k-N-1|k} \tag{3.32}$$

$$e_{j+1|k} = Ae_{j|k} - \hat{w}_{j|k} \qquad ,j = k-N, \ldots, k-1 \tag{3.33}$$

with $\bar{e}_{k-N-1} = x_{k-N-1} - \bar{x}_{k-N-1}$. This formulation is very similar to the MPC formulation given in (3.25)-(3.26) and illustrates why MHE can be seen as the state estimation counterpart to the MPC problem. However there are still significant differences:

- In the MPC formulation $u_{k|k}$ is needed for the implementation of the controller. It is the first value in the "optimization–sequence". In the MHE problem the actual state estimate is the last value in the "optimization–sequence". The following section will show how this difference leads to significant differences of the "solution" structure of the MHE and MPC formulations, if a dynamic programming approach is used to calculate the next estimates/controls.

[3]The consideration of stability under real noise influence would equal the question of the robustness of MPC strategies under model mismatch and noise influence. This is a much more difficult problem.

- For the MHE problem the "final–state–penalty" term of the MPC formulation $x^T_{k+N|k} Q_N x_{k+N|k}$ is moved to the beginning of the horizon as $\hat{w}^T_{k-N-1|k} Q^{-1}_{-N} \hat{w}_{k-N-1|k}$. Additional a final error term $e^T_{k|k} C^T R^{-1} C e_{k|k}$ shows up.

- There is no constant "unchangeable" term corresponding to $x^T_{k|k} * Q * x_{k|k}$ of the MPC formulation, in the MHE formulation.

3.3.2 Dynamic Programming Solution for the Unconstrained MHE

It is possible to calculate an analytic solution for the unconstrained linear MHE via Dynamic Programming (DP) [1, 3]. The optimization problems given by equations (3.31)-(3.32) and equations (3.25)-(3.26) can be seen as multi-stage problems. Multi-stage stands here for the fact that the state x_j depends only on the previous state x_{k-1} and input u_{k-1}. This forward coupling allows to solve the optimization problem stage-wise backwards. The first step is to assume that all optimization variables except the last one (here the inputs u_j or disturbances $\hat{w}_j$) are already known. The remaining one variable optimization problem is solved in dependence of the previous one. Then the problem is backed up one step into the past and the whole methods starts new. This is often referred to as the Bellman method [1] for the solution of multi stage optimization problems. A graphical interpretation of this approach for the MPC case is given in Figure 3.2. During this section the dynamic programming solution for

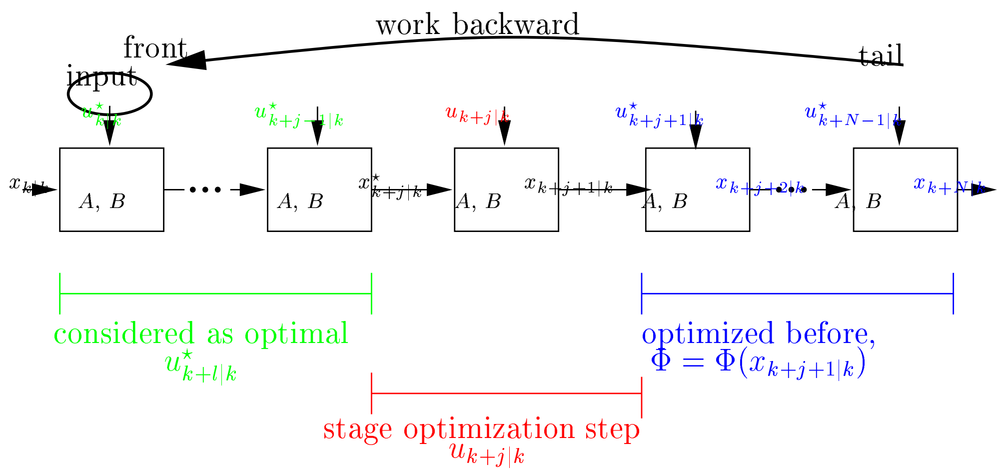

Figure 3.2: Dynamic programming approach for MPC.

both, the finite horizon MPC problem and the MHE problem are given. The major difference will be that for the MPC solution only the "first" optimization variable $u_{k|k}$ is needed. This value is the result of working from the tail $j = k + N - 1$ to the front $j = k$. The values $u_{j|k},\ j = k+1, \ldots, j = k+N-1$ are not needed in order to calculate the new control move u_k. The MHE set up given in equations (3.31)-(3.32) requires as next estimate the value $\hat{w}_{k|k}$. However the dynamic programming method used in the MPC formulation gives only the value of $\hat{w}_{k-N-1|k}$, so that an additional backup to $\hat{w}_{k|k}$ is necessary (see Figure 3.3 for a graphical explanation).

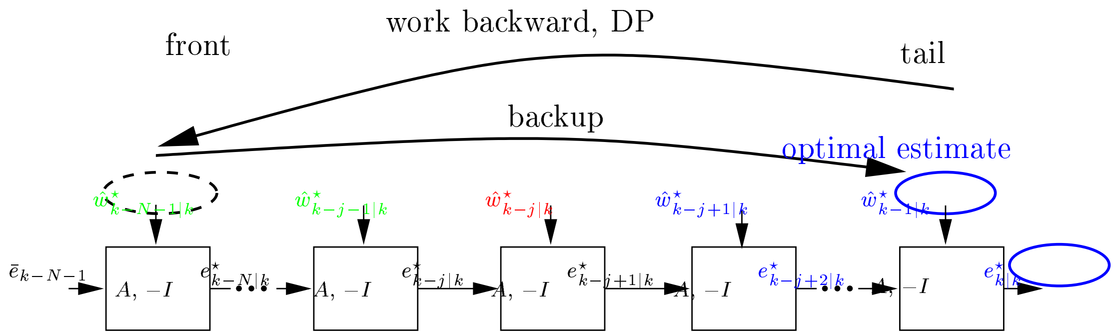

Figure 3.3: Dynamic programming approach for MHE.

Motivation: A Dynamic Programming Solution for the Linear Unconstrained MPC Problem

The derivations given here are based on the work of Rawlings [39]. They are included in order to allow for a better understanding of the DP approach for the MHE. They also allow for a comparison between the MPC and MHE solution.

On of the key steps of the dynamic programming approach for multistage optimization problems[4] is, to consider the optimization problem as solved in all optimization variables except the last one. This means that the inputs $u_{k+j|k}$ up to time $N-2$ are assumed optimal and fixed $\{u_k^\star|k, \ldots u_{k+N-2|k}^\star\}$. The remaining cost is then minimized with respect to $u_{k+N-1|k}$. This minimization problem is given by:

$$\min_{u_{k+N-1|k}} \Phi_{k+N-1|k} : \quad \Phi_{k+N-1|k} = x^{\star T}_{k+N-1|k} \tilde{Q} x^\star_{k+N-1|k} + u^T_{k+N-1|k} \tilde{R} u_{k+N-1|k} + x^T_{k+N|k} \tilde{Q}_N x_{k+N|k} \quad (3.34)$$

subject to

$$x_{k+N|k} = A x^\star_{k+N-1|k} + B u_{k+N-1|k} \quad (3.35)$$

[4]The cost function consists of stage cost functions $L(x_{j|k}, u_{j|k}) = x^T_{j|k} \tilde{Q} x_{j|k} + u^T_{j|k} \tilde{R} u_{j|k}$. The state $x_{j|k}$ at every time only depends on the previous state and input: $x_{j|k} = A x_{j-1|k} + B u_{j-1|k}$

Here the index $k+N-1|k$ is used to express that $\Phi_{k+N-1|k}$ is the cost for moving from $k+N-1$ to $k+N$. The equality constraints (3.35) can be removed by substitution into Φ. This leads to:

$$\Phi_{k+N-1|k} = {x^\star}^T_{k+N-1|k}\left(A^T\tilde{Q}_N A + \tilde{Q}\right)x^\star_{k+N-1|k} + u^T_{k+N-1|k}\left(B^T\tilde{Q}_N B + \tilde{R}\right)u_{k+N-1|k} + 2{x^\star}^T_{k+N-1|k}B^T\tilde{Q}_N A u_{k+N-1|k}$$

Since the term with $u_{k+N-1|k}$ is positive definite and quadratic, a unique maxima exists and can be calculated by:

$$\left.\frac{\partial\Phi_{k+N-1|k}}{\partial u_{k+N-1|k}}\right|_{u^\star_{k+N-1|k}} = 0$$
$$= 2B^T\tilde{Q}_N A{x^\star}_{k+N-1|k} + 2\left(B^T\tilde{Q}_N B + \tilde{R}\right)u^\star_{k+N-1|k}$$

This leads to the following optimal input $u^\star_{k+N-1|k}$ in dependence of ${x^\star}^T_{k+N-1|k}$.

$$\begin{aligned} u^\star_{k+N-1|k} &= -\left(B^T\tilde{Q}_N B + \tilde{R}\right)^{-1}B^T\tilde{Q}_N A{x^\star}_{k+N-1|k} \\ &= K_{N-1}{x^\star}_{k+N-1|k} \end{aligned} \tag{3.36}$$

Here K_{N-1} instead of $K_{k-1|k}$ is used since K does not depend on k, it is time invariant and fixed.

The optimal cost for $\Phi_{k+N-1|k}$ becomes

$$\begin{aligned} \Phi^\star_{k+N-1|k} &= {x^\star}^T_{k+N-1|k}\left(\tilde{Q} + A^T\left(\tilde{Q}_N - \tilde{Q}_N B(\tilde{R} + B^T\tilde{Q}_N B)^{-1}B^T\tilde{Q}_N\right)A\right){x^\star}_{k+N-1|k} \\ &= {x^\star}^T_{k+N-1|k}\Pi_{N-1}{x^\star}_{k+N-1|k} \end{aligned} \tag{3.37}$$

The cost is quadratic in ${x^\star}_{k+N-1|k}$ and $u^\star_{k+N-1|k}$ depends linearly via K_{N-1} on ${x^\star}^T_{k+N-1|k}$.

For the first stage it was assumed that ${x^\star}_{k+N-2|k}$ was known and therefore also the optimal sequence $\{u^\star_{k|k}, \ldots u^\star_{k+N-2|k}\}$ was known. In order to go on with the optimization the stage cost-function for $\Psi_{k+N-2|k}$ is expressed using the just calculated ${x^\star}_{k+N-1|k}$. This results in a new one variable optimization problem for ${x^\star}_{k+N-2|k}$ if the inputs $u_{k+j|k},\ j = 0, \ldots, N-3$ are fixed, to their optimal value $\{u^\star_{k|k}, \ldots u^\star_{k+N-3|k}\}$. The optimization problem for $u^\star_{k+N-2|k}$ or ${x^\star}_{k+N-2|k}$ respectively becomes

$$\min_{u_{k+N-2|k}} \Phi_{k+N-2|k}: \quad \Phi_{k+N-2|k} = {x^\star}^T_{k+N-2|k}\tilde{Q}x^\star_{k+N-2|k} + u^T_{k+N-2|k}\tilde{R}u_{k+N-2|k} + x^T_{k+N|k}\Pi_{N-1}x_{k+N|k}$$

with

$$x_{k+N|k} = Ax_{k+N-1|k} + Bu_{k+N-1|k}$$

This formulation has the same structure as the problem given by equations (3.34)-(3.35). Replacing $\tilde{Q}_N$ with Π_{N-1} in (3.36), yields the following:

$$\begin{aligned} u^{\star}_{k+N-2|k} &= \underbrace{-\left(B^T\Pi_{N-1}B + \tilde{R}\right)^{-1} B^T\Pi_{N-1}A}_{K_{N-2}} x^{\star}{}_{k+N-2|k} \\ \Phi^{\star}_{k+N-2|k} &= x^{\star T}{}_{k+N-2|k} \\ &\quad \underbrace{\left(\tilde{Q} + A^T\left(\Pi_{N-1} - \Pi_{N-1}B(\tilde{R} + B^T\tilde{Q}_N B)^{-1}B^T\Pi_{N-1}\right)A\right)}_{\Pi_{N-2}} x^{\star}{}_{k+N-2|k} \end{aligned}$$

The whole process can now be continued until $u^{\star}_{k|k}$ is reached. This leads to the so called discrete time Riccati equation (DRE) for the calculation of Π_j.

$$\Pi_{j-1} = \tilde{Q} + A^T\left(\Pi_j - \Pi_j B(\tilde{R} + B^T\tilde{Q}_N B)^{-1}B^T\Pi_j\right)A \quad j = N, \ldots, 1 \tag{3.38}$$

with the initial value

$$\Pi_N = \tilde{Q}_N \tag{3.39}$$

The solution of the optimal finite horizon LQ/MPC problem can now be given in terms of the Riccati matrix Π_j:

$$u_k = u^{\star}_{k|k} \tag{3.40}$$

$$= K_0 x_k \tag{3.41}$$

with K_0 as

$$K_0 = -\left(B^T\Pi_1 B + \tilde{R}\right)^{-1} B^T\Pi_1 A \tag{3.42}$$

and Φ_k

$$\Phi_k = x_k^T\Pi_0 x_k \tag{3.43}$$

Dynamic Programming Solution for Linear Unconstrained MHE

Motivated by the DP solution of the MPC problem, it is possible to derive an explicit solution of the MHE. The results will be similar, however since the optimal estimate

$e_{k|k}$ (respectively $\hat{x}_{k|k}$) is the "last" optimization variable, additional steps will be necessary (see Figure 3.3 and Figure 3.2).

Starting point is the optimization problem given in (3.31)-(3.32). Similar to the DP solution for the MPC problem the state errors $\hat{w}_{k-j|k}$, $j = N+1, \ldots, 2$ are assumed to be fixed and optimal. $\{\hat{w}^\star_{k-N-1|k}, \ldots, \hat{w}^\star_{k-2|k}\}$. With this the optimization problem reduces to

$$\min_{\hat{w}_{k-1|k}} \Psi_{k-1|k} : \quad \Psi_{k-1|k} = \hat{w}^T_{k-1|k} Q^{-1} \hat{w}_{k-1|k} + {e^\star_{k-1|k}}^T C^T R^{-1} C e^\star_{k-1|k} + e^T_{k|k} C^T R^{-1} C e_{k|k}$$

with $e_{k|k} = Ae^\star_{k-1|k} - \hat{w}_{k-1|k}$.
Substituting this into the equation for $\Psi_{k-1|k}$ gives:

$$\Psi_{k-1|k} = {e^\star_{k-1|k}}^T C^T R^{-1} C e^\star_{k-1|k} + \hat{w}^T_{k-1|k} \left(Q^{-1} + \Upsilon_0\right) \hat{w}_{k-1|k} + {e^\star_{k-1|k}}^T A^T \Upsilon_0 A e^\star_{k-1|k} - 2\hat{w}^T_{k-1|k} \Upsilon_0 A e^\star_{k-1|k} \quad (3.44)$$

with $\Upsilon_0 = C^T R^{-1} C$. The matrix Υ_j will play the role of the stage-cost matrix Π_j. The optimal $\hat{w}^\star_{k-1|k}$ can now be be calculated by

$$\left.\frac{\partial \Psi_{k-1|k}}{\partial \hat{w}_{k-1|k}}\right|_{\hat{w}^\star_{k-1|k}} = 0 = 2\left(Q^{-1} + \Upsilon_0\right) \hat{w}^\star_{k-1|k} - 2\Upsilon_0 A e^\star_{k-1|k}$$

Similar to the MPC formulation the resulting optimal value for $\hat{w}^\star_{k-1|k}$ depends linearly on $e^\star_{k-1|k}$.

$$\hat{w}^\star_{k-1|k} = \left(Q^{-1} + \Upsilon_0\right)^{-1} \Upsilon_0 A e^\star_{k-1|k} = L_{-1} e^\star_{k-1|k}$$

The optimal cost $\Psi^\star_{k-1|k}$ is given by

$$\Psi^\star_{k-1|k} = {e^\star_{k-1|k}}^T \left(L^T_{-1} Q^{-1} L_{-1} + (A - L_{-1})^T \Upsilon_0 (A - L_{-1}) + C^T R^{-1} C\right) e^\star_{k-1|k} = {e^\star_{k-1|k}}^T \underbrace{\left(C^T R^{-1} C + A^T \left(\Upsilon_0 - \Upsilon_0 \left(Q^{-1} + \Upsilon_0\right)^{-1} \Upsilon_0\right) A\right)}_{\Upsilon_{-1}} e^\star_{k-1|k}$$

Changing now to the next stage cost $\Upsilon_{k-2|K}$, using similar arguments as before yields the following minimization which must be solved in order to get $\hat{w}^\star_{k-2|k}$:

$$\min_{\hat{w}^\star_{k-2|k}} \Psi_{k-2|k} : \quad \Psi_{k-1|k} = \hat{w}^T_{k-2|k} Q^{-1} \hat{w}_{k-2|k} + {e^\star_{k-2|k}}^T C^T R^{-1} C e^\star_{k-2|k} + e^T_{k-1|k} \Upsilon_{-1} e_{k-1|k}$$

subject to

$$e_{k-1|k} = Ae^{\star}_{k-2|k} - \hat{w}_{k-2|k}$$

Comparison with the similar formulation (3.44) of the MPC problem shows that the derived formulation has the same structure. This allows to calculate all $\hat{w}_{k-j|k}$, except the "last" $\hat{w}_{k-N-1}$ using the same equations as used in the MPC formulation. The last $\hat{w}$ must be calculated separately, since the coupling matrix Q^{-1}_{-N} does not equal Q^{-1} and also the initial value $\bar{e}_{k-N-1}$ is unspecified.

After applying several times the from the MPC formulation known algorithm the stage cost function for $\hat{w}_{k-N-1|k}$ becomes:

$$\begin{aligned}\Psi_k &= \Psi_{k-N|k} \\ &= \hat{w}^T_{k-N-1|k} Q^{-1}_{-N} \hat{w}_{k-N-1|k} + e^T_{k-N|k} \Upsilon_{-N} e_{k-N|k}\end{aligned}$$

where $e_{k-N|k}$ is given by

$$e_{k-N|k} = A\bar{e}_{k-N-1} - \hat{w}_{k-N-1|k}$$

so that after substituting this in Ψ_k becomes

$$\begin{aligned}\Psi_k = \hat{w}^T_{k-N-1|k} \left(Q^{-1}_{-N} + \Upsilon_{-N}\right) \hat{w}_{k-N-1|k} - 2\hat{w}_{k-N-1|k} \Upsilon_{-N} A\bar{e}_{k-N-1} + \\ \bar{e}^T_{k-N-1} A^T \Upsilon_{-N} A\bar{e}_{k-N-1}\end{aligned}$$

Minimization with respect to $\hat{w}_{k-N-1|k}$ leads to

$$\begin{aligned}\hat{w}^{\star}_{k-N-1|k} &= \left(Q^{-1}_{-N} + \Upsilon_{-N}\right)^{-1} \Upsilon_{-N} A\bar{e}_{k-N-1} \\ &= L_{-N-1}\bar{e}_{k-N-1} \end{aligned} \tag{3.45}$$

The optimal cost becomes

$$\begin{aligned}\Psi^{\star}_k(\bar{e}_{k-N-1}) &= \bar{e}^T_{k-N-1} A^T \left(\Upsilon_{-N} - \Upsilon_{-N}\left(Q^{-1}_N + \Upsilon_{-N}\right)^{-1} \Upsilon_{-N}\right) A\bar{e}_{k-N-1} \\ &= \bar{e}^T_{k-N-1} \Upsilon_{-N-1} \bar{e}_{k-N-1}\end{aligned} \tag{3.46}$$

To summarize, the following iteration can be used for the estimates $\hat{w}_{k-j|k}$, $j = 1, \ldots, N+1$.

$$\begin{aligned}\hat{w}^{\star}_{k-j|k} &= L_{-j} e^{\star}_{k-j|k} \quad j = 1, \ldots, N \\ \hat{w}^{\star}_{k-N-1|k} &= L_{-N-1}\bar{e}_{k-N-1}\end{aligned} \tag{3.47}$$

where L_{-j} can be calculated as follows

$$\begin{aligned} L_{-j} &= \left(Q^{-1} + \Upsilon_{-j+1}\right)^{-1} \Upsilon_{-j+1} A \quad j = 1, \ldots, N \\ L_{-N-1} &= \left(Q_{-N}^{-1} + \Upsilon_{-N}\right)^{-1} \Upsilon_{-N} A \end{aligned} \tag{3.48}$$

and Υ_{-j} is given by the following Riccati iteration with modified $N = 1$'th value

$$\begin{aligned} \Upsilon_{-j} &= C^T R^{-1} C + A^T \left(\Upsilon_{-j+1} - \Upsilon_{-j+1} \left(Q^{-1} + \Upsilon_{-j+1}\right)^{-1} \Upsilon_{-j+1}\right) A \quad j = 1, \ldots, N \\ \Upsilon_{-N-1} &= A^T \left(\Upsilon_{-N} - \Upsilon_{-N} \left(Q_{-N}^{-1} + \Upsilon_{-N}\right)^{-1} \Upsilon_{-N}\right) A \end{aligned} \tag{3.49}$$

$$\Upsilon_0 = C^T R^{-1} C \tag{3.50}$$

The optimal cost and the cost to go can be calculated using Υ_{-j}

$$\begin{aligned} \Psi^\star_{k-j|k} &= e^\star_{k-j|k} \Upsilon^T_{-j} e^\star_{k-j|k} \quad j = 1, \ldots, N \\ \Psi^\star_k &= \bar{e}^T_{k-N-1} \Upsilon_{-N-1} \bar{e}_{k-N-1} \end{aligned} \tag{3.51}$$

In order to get an equation for $e^\star_{k|k}$ the following back-up scheme starting from $\bar{e}_{k-N-1}$ is used

$$\begin{aligned} e^\star_{k-N|k} &= A\bar{e}_{k-N-1} - \hat{w}^\star_{k-N-1|k} \\ &= \left(A - L_{-N-1}\right) \bar{e}_{k-N-1} \end{aligned} \tag{3.52}$$

$$\begin{aligned} e^\star_{k-N+1|k} &= Ae^\star_{k-N|k} - \hat{w}^\star_{k-N|k} \\ &= \left(A - L_{-N}\right) e^\star_{k-N|k} \\ &= \left(A - L_{-N}\right)\left(A - L_{-N-1}\right) \bar{e}_{k-N-1} \end{aligned} \tag{3.53}$$

$$\vdots$$

$$e_k = e_{k|k} = \left(A - L_{-1}\right) \cdots \left(A - L_{-N-1}\right) \bar{e}_{k-N-1} \tag{3.54}$$

With this the equation connecting $e_{k|k}\star$ to $\bar{e}_{k-N-1}$ is found. In order to derive sufficient conditions for the nominal stability of the given MHE scheme, an choice for the update formulation of $\bar{x}^\star_{k-N-a}$ must be made.

Remarks: Similar results could be derived by using the concept of duality for linear systems [17]. The major obstacle here is the difference in the formulation of the MPC problem (3.25)-(3.26) and the MHE problem. Major differences in the formulations are that for the MHE formulation two "end penalty" terms show u, one at the end of the horizon (which runs backward), $\hat{w}^T_{k-N-1|k} Q^{-1}_{-N}\, \hat{w}_{k-N-1|k}$ and one at the beginning, $\hat{v}^T_{k|k} R^{-1} \hat{v}_{k|k}$. This can also be seen in the more comparable error formulation (3.28)-(3.29). Therefore a direct application of the duality concept as used for the comparison of the infinite horizon LQR and KF seems to be more difficult. △

3.3.3 Different Update Schemes

This section discusses the different choices for the update of the initial state value $\bar{x}_{k-N-1}$. The transfer of "information", which is available in the form of previous estimates, to the actual estimation via $\bar{x}_{k-N-1}$ seems to be viable for the stability of the estimator. There are many different possibilities to handle this "information" transfer. Two useful and also limiting cases are:

- the usage of $\hat{x}^{\star}_{k-N-1|k-N-1}$, the "filtered estimate" calculated $k-N-1$ steps in the past.
- the usage of the last smoothed value $\hat{x}^{\star}_{k-N-1|k-1}$

It is also possible to use other values, often smoothed ones like $\hat{x}^{\star}_{k-N-1|k-j}$. However, the major advantages and disadvantages of the different update schemes can be explored with the two presented methods.

The Filtered Estimate $\hat{x}^{\star}_{k-N-1|k-N-1}$ Update

The major argument for the usage of $\hat{x}^{\star}_{k-N-1|k-N-1}$ lays in the fact, that no input information y_j is over weighted during the estimation. The y_j used during the estimation of $\hat{x}^{\star}_{k-N-1|k-N-1}$ lay before the current end of the horizon, $k-N$. This guarantees that no measurement information is used twice. If in contrary a smoothed update like $\hat{x}^{\star}_{k-N-1|k-1}$ is used, input information is used more then once. This is not favorable from a probabilistic point of view. However one disadvantage of the usage of $\hat{x}^{\star}_{k-N-1|k-N-1}$ together with a constant $Q_{-N|k}$ lays in the fact that an "estimation" cycling like shown in Figure 3.4 can occur. This effect can be explained by dividing

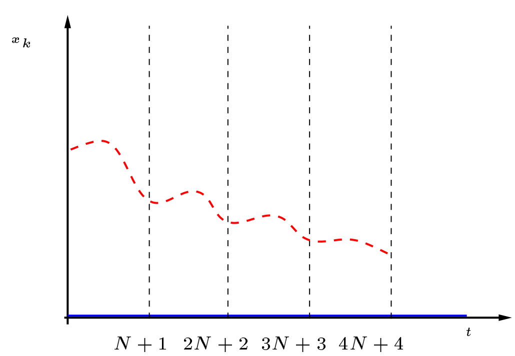

Figure 3.4: Cycling due to update every N+1 steps.

(in the nominal case) the resulting estimation algorithm in N "independently" and by one time step shifted running estimators with a sampling time of N+1 time-steps. Once these estimators are started up with the first N initial estimates (generated for

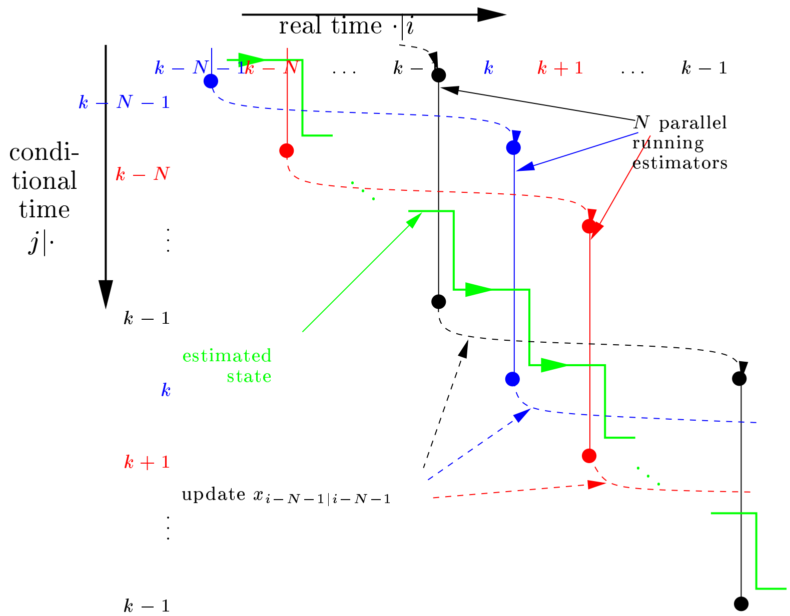

Figure 3.5: Cycling explanation for filtered update.

example using a batch estimator), only estimates calculated by the same estimator are coupled (see Picture 3.5). This effect will also show up in the in section 3.3.4 derived stability condition for an MHE with filtered update. Figure 3.6 gives an example simulation for this behavior. The considered system was taken from Muske and Rawlings [35]. It consists out of an unstable system with one state. The system equations are given by

$$\begin{aligned} x_{k+1} &= 1.1x_k \\ y_k &= x_k \end{aligned}$$

The estimator parameters were chosen as follows: $Q^{-1} = Q^{-1}_{-N} = 1$, $R^{-1} = 1/100$, horizon length $N = 13$. The given system is observable and additional $(A, Q^{1/2})$ is controllable. The real system state was chosen to be 0 at the startup, what translates to $x_k = 0 \; \forall \; k \geq 0$. As nominal startup disturbance an value of $\bar{x}_0 = 2$ was taken.
Figure 3.6 shows clearly the cycling of the state during the estimation, introduced by the filtered update scheme.

The Smoothed Estimate $\hat{x}^{\star}_{k-N-1|k-1}$ Update

This update method uses the one time-step before calculated smoothed state estimate $\hat{x}^{\star}_{k-N-1|k-1}$. Due to the "shorter" coupling period this update scheme has no artificial estimator "cycling". However, there is no direct probabilistic motivation for the use

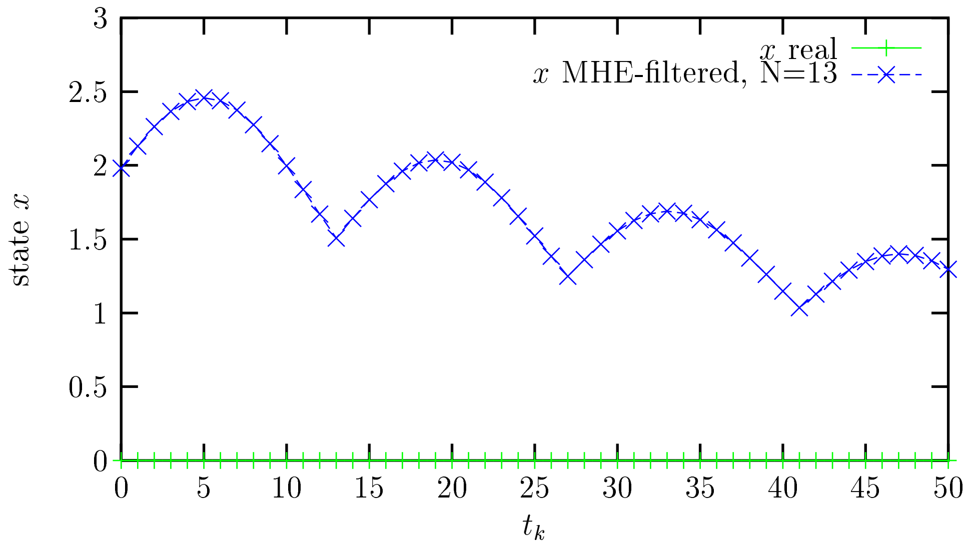

Figure 3.6: Example for cycling effect, $A = 1.1$, $C = 1$, $Q^{-1} = Q_{-N}^{-1} = 1$, $R^{-1} = 1/100$.

of this value. The closer connection between the last estimate and the actual estimate allows on to derive similar stability results as for the MPC with finite horizon. For this connections see section 3.3.4. A graphical illustration of the smoothed update method is given in Figure 3.7.
Figure 3.8 shows the result of MHE with smoothing update on the same system and with the same parameters as used for the MHE example with filtered update. It is clearly to see that no "artificial" estimator cycling is introduced. Additionally it can be seen that the smoothed update leads to an faster convergence of the estimation error to 0.

3.3.4 Stability of the Unconstrained MHE

In this section sufficient stability conditions for the unconstrained MHE with filtered and smoothed update are derived. Since the filtered update uses an N-step old estimate as initial value , where as the MHE with smoothed update uses a 1-step old estimate, the stability conditions will differ slightly.

Filtered Update

The stability condition for the filtered update is based on equation (3.54). This equation gives a direct connection between $N + 1$ steps apart estimates. As shown in section 3.3.3, the MHE with filtered update can be seen as N parallel running

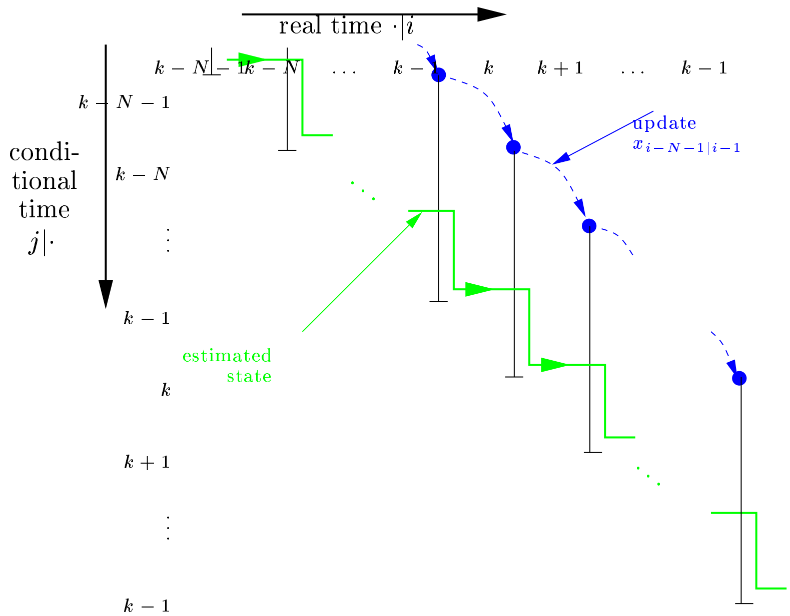

Figure 3.7: Smoothed update, no cycling.

estimators. Since these estimators are “not” coupled among themselves the stability proof must be based adjacent estimates coming from the same estimator. In order to guarantee stability the dynamic matrix given by equation (3.54) that couples two $N+1$ steps apart estimation errors must be stable:

$$\lambda_{\text{max}} = \max|\lambda_i\left((A-L_{-1})\cdots(A-L_{-N-1})\right)| < 1 \tag{3.55}$$

This leads to the following Theorem

Theorem 3.3 Stability Condition for MHE with Q_{-N} =const and Filtered Initial Update *The unconstrained linear MHE as given in (3.28)-(3.29) with $\bar{x}_{k-N-1} = \hat{x}^{\star}_{k-N-1|k-N-1}$ and fixed values for N, $Q^{-1} > 0$, $Q^{-1}_{-N} > 0$, $R^{-1} > 0$ is asymptotically stable if,*
1) $|\lambda\left((A-L_{-1})\cdots(A-L_{-N-1})\right)|_{max} < 1$,
2) the N initial estimates $\hat{x}_{j|k}$, $j = 1, \dots N$ are finite.
$|\lambda(\cdot)|_{max}$ corresponds to the maximum absolute eigenvalue.

Proof: The proof follows directly from equation (3.54). The use of $\hat{x}^{\star}_{k-N-1|k-N-1}$ as $\bar{x}_{k-N-1}$ leads to $\bar{e}_{k-N-1} = \bar{e}^{\star}_{k-N-1|k-N-1}$, so that the $N+1$ step error propagation becomes

$$e_k = e_{k|k} \quad = \underbrace{(A-L_{-1})\cdots(A-L_{-N-1})}_{\Lambda}\bar{e}^{\star}_{k-N-1|k-N-1}$$

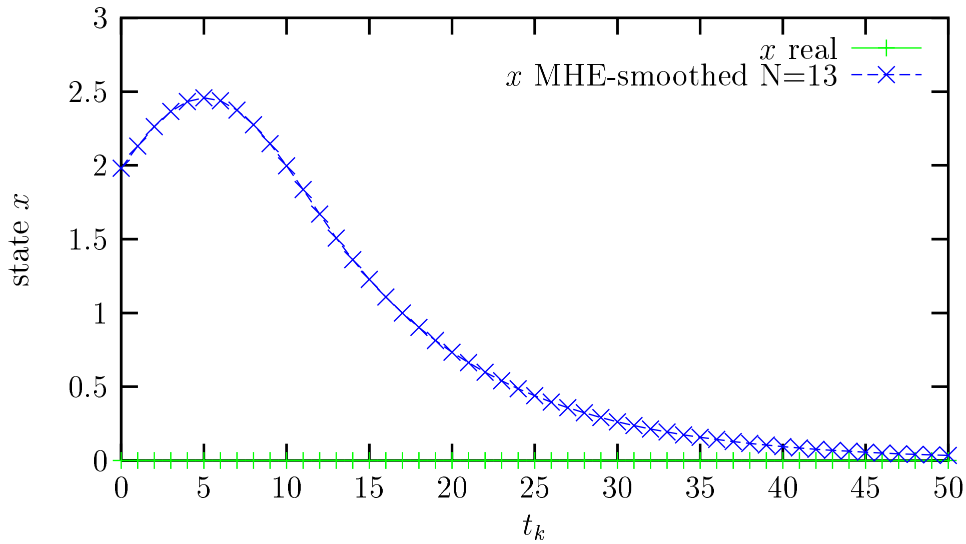

Figure 3.8: Example for smoothed update, no cycling, $A = 1.1$, $C = 1$, $Q^{-1} = Q_{-N}^{-1} = 1$, $R^{-1} = 1/100$.

Under the assumptions given in the theorem the error dynamic matrix Λ has eigenvalues strictly inside of the unit circle. As result of this the $N+1$ step error propagation is stable. Since k is not specified and all initial estimates are finite, the error converges for all N independent error sequences to 0. Hence the MHE with filtered update is asymptotically stable. This concludes the proof. $\square$

The important part of this proof is that it was possible to explicitly calculate the estimation error dynamic as given in (3.54). The proof can be compared to the fact that every finite horizon unconstrained MPC is stable, as long as $(A + BK_0)$ is stable. The condition derived here is more complicated, since a "backup" to the last optimization variable $\hat{w}_{k|k}$ is necessary.

Smoothed Update

The stability condition for the smoothed update grounds on the same basis as the one for the MHE with filtered initial update. However in order to guarantee stability of the estimation error the connection between the smoothed errors $e^{\star}_{k-N|k}$ and $e^{\star}_{k-N-1|k-1}$ must be examined.

Theorem 3.4 Stability Condition for MHE with Q_{-N} and Smoothed Update *The unconstrained linear MHE as given in (3.28)-(3.29) with $\bar{x}_{k-N-1} = \hat{x}^{\star}_{k-N-1|k-1}$ and fixed values for N, $Q^{-1} > 0$, $Q^{-1}_{-N} > 0$, $R^{-1} > 0$ is asymptotically stable if,*

1) $|\lambda(A - L_{-1})|_{max} < 1$,
2) the initial N *estimates* $\hat{x}_{j|k}$, $j = 1, \dots N$ *are finite.*
$|\lambda(\cdot)|_{max}$ *corresponds to the maximum absolute eigenvalue.*

Proof: The proof is similar to the one given for the filtered update. $\bar{x}_{k-N-1} = \hat{x}^{\star}_{k-N-1|k-1}$ implies that $\bar{e}_{k-N-1} = \bar{e}^{\star}_{k-N-1|k-1}$, which results in the following error propagation equation:

$$e^{\star}_{k-N|k} = \underbrace{(A - L_{-N-1})}_{\Lambda} \bar{e}^{\star}_{k-N-1|k-1}$$

Due to assumption 1) and (3.52) we know that Λ for the estimation error has only eigenvalues inside the unit circle, and by this $e^{\star}_{k-N|k}$ converges to 0 and is asymptotically stable. Since by equation (3.53)- (3.54) all other estimation errors $e^{\star}_{k-j|k}$, especially $e^{\star}_{k|k} = e_k$, depend on $e^{\star}_{k-N|k}$, they also converge to 0. This concludes the proof. □

This stability condition can be seen as the direct takeover of the stability condition for linear MPC with finite horizon-length, that $(A + BK_0)$ has to be stable in order to guarantee asymptotic stability.

3.3.5 Examples

The during the previous section derived stability conditions allow to analyze the stability for fixed Q^{-1}_{-N}. The here given examples will verify the given theorems for the smoothed and filtered update. One of the major results will be, that the tuning of a MHE filter with fixed Q^{-1}_{-N} is nonintuitive. Example 2 will show, that for different horizon length the resulting MHE can be stable for one horizon and unstable for the next higher one, and so on. For this reason the given test seems to be very important, since it allows to test a priory if a chosen set of parameters Q^{-1}, R^{-1}, Q^{-1}_{-N} and N leads to a nominal stable estimator. The given examples are on purpose very simple and of small order, in order to allow insight in the behavior of the estimator.

Example 1., Stability for Different Initial Weights

The following example taken from [35] is used to show the influence of different initial weights. The chosen system is unstable, the initial estimate is set to $\bar{x}_0 = 2$, the state of the “real” nominal system is $x_j = 0$.

$$\begin{aligned} x_{k+1} &= 1.1 x_k \\ y_k &= x_k \end{aligned}$$

The horizon length is fixed to $N = 12$ for both, the filtered and smoothed estimator. It is obvious that the given system is observable. Under the initial estimate $\bar{x}_0$ the estimation problem can be seen as the recovery from a wrong initial value. The weights Q^{-1} and R^{-1} are chosen to

$$Q^{-1} = 1, \quad R^{-1} = 1/100$$

which might correspond for the real problem under noise influence to small state w but large measurement noise v, since $Q >> R$.

Figure 3.9 and 3.10 correspond to the following choice of initial weights:

$$\begin{array}{rl} \text{Picture 3.9:} & Q_{-N}^{-1} = 2 \\ \text{Picture 3.10:} & Q_{-N}^{-1} = 0.1 \end{array}$$

As can be in Figure 3.9 the filtered estimator is unstable for $Q_{-N}^{-1} = 2$, whereas

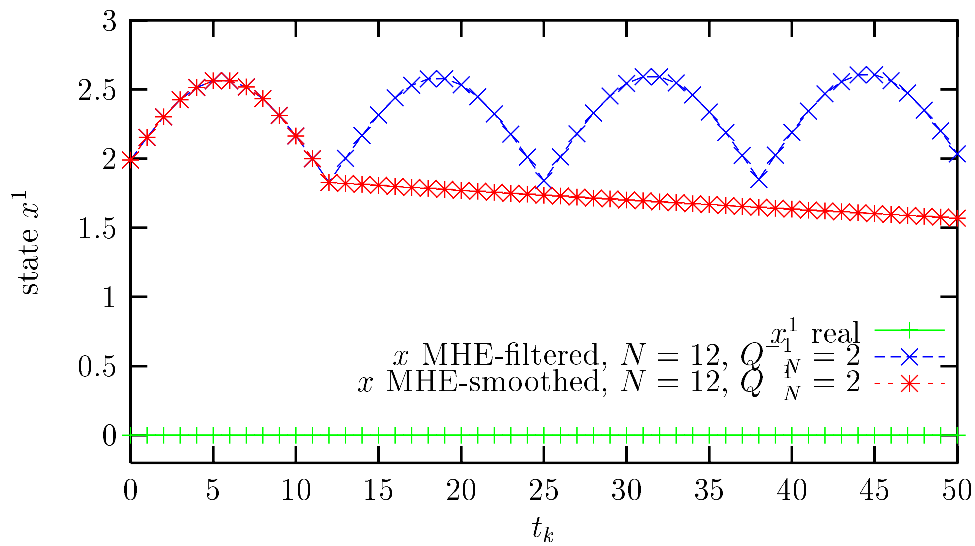

Figure 3.9: Example 1.: Stability in dependence of Q_{-N}, $Q_{-N}^{-1} = 2$, $N = 12$.

the smoothed one is stable for this value. The maximum amplitude of the filtered estimator is strictly increasing, whereas the smoothed estimate slowly decreases. This fact can be confirmed by calculating the resulting eigenvalues of the matrix Λ as given in Theorem 3.3 and 3.4.

$$\begin{aligned} |\lambda|^{\text{filtered}} &= 1.01 \\ |\lambda|^{\text{smoothed}} &= 0.99 \end{aligned}$$

The two graphs coincide for the first N steps, since the same batch estimator is used for the startup phase. This will be also seen in the later following examples. For the

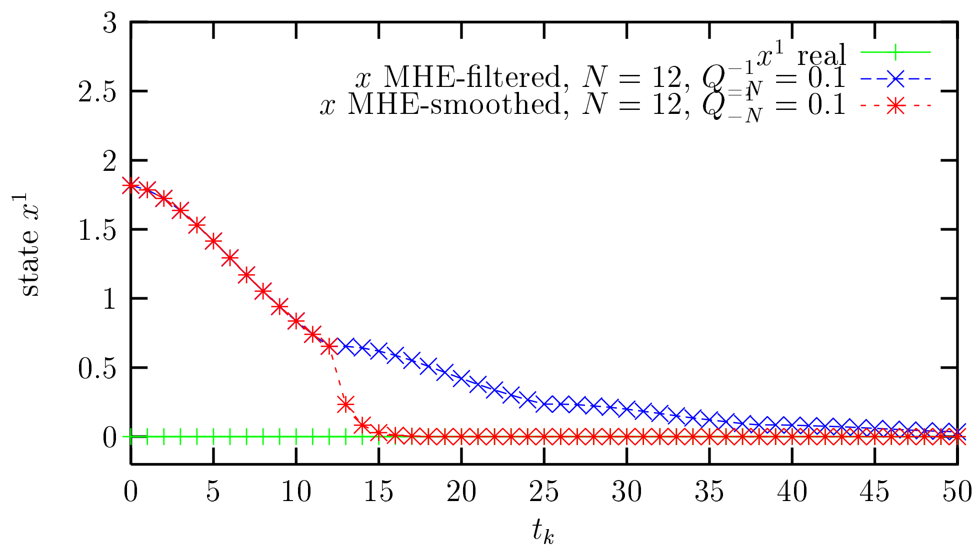

Figure 3.10: Example 1.: Stability in dependence of Q_{-N}, $Q_{-N}^{-1} = 0.1$ $N = 12$.

choice of $Q_{-N}^{-1} = 0.1$ both estimators are stable, however the smoothed one converges faster to the true value 0. The eigenvalues of Λ for this case can be calculated to

$$\begin{aligned} |\lambda|^{\text{filtered}} &= 0.36 \\ |\lambda|^{\text{smoothed}} &= 0.35 \end{aligned}$$

Both pictures clearly show the cycling effect for the filtered estimate. This is even obvious for the stable and relatively "fast" converging case with $Q_{-N}^{-1} = 0.1$.

Example 2., a Second Order System

The second example is taken from Muske and Rawlings [40].

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} \frac{5}{3} & -\frac{2}{3} \\ 1 & 0 \end{bmatrix} x_k \\ y_k &= \begin{bmatrix} \frac{-2}{3} & 1 \end{bmatrix} x_k \end{aligned}$$

This system is observable and has the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = \frac{2}{3}$. The weight matrices are chosen to

$$Q^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad R^{-1} = 1/100,$$

the horizon is fixed to 3. The following two initial weight matrices where used:

$$\text{for Figure 3.11} \quad Q_{-N}^{-1} = \begin{bmatrix} 300 & 0 \\ 0 & 300 \end{bmatrix}$$

$$\text{for Figure 3.12} \quad Q_{-N}^{-1} = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.00 \end{bmatrix}$$

As can be seen in Figure 3.11, the first choice of Q_{-N}^{-1} does not lead to a stable estimator. For simplicity only the first state x^1 is shown. The MHE with filtered estimated is instable, whereas the MHE with smoothed estimate stays at the initial estimate 4. The for Λ resulting eigenvalues for both initial estimates are given in

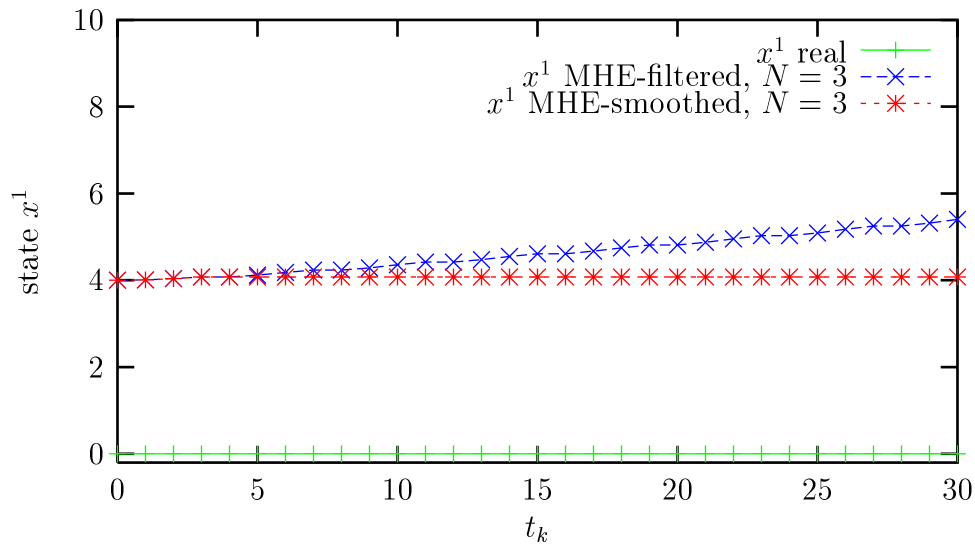

Figure 3.11: Example 2.: Stability in dependence of Q_{-N}^{-1}, $N = 3$.

Table 3.3.5. As can be seen in Figure 3.12 and from Table 3.3.5 the second choice

Q_{-N}^{-1}	filtered		smoothed	
	λ_1	λ_2	λ_1	λ_2
$Q_{-N}^{-1} = \left[\begin{smallmatrix} 300 & 0 \\ 0 & 300 \end{smallmatrix}\right]$	1.044	0.180	1.000	0.666
$Q_{-N}^{-1} = \left[\begin{smallmatrix} 0.001 & 0 \\ 0 & 0.00 \end{smallmatrix}\right]$	0.52	0.006	0.422	0.029

Table 3.1: Eigenvalues Λ example 2 in dependence of Q_{-N}^{-1}

for Q_{-N}^{-1} results in nominal asymptotically stable estimators.

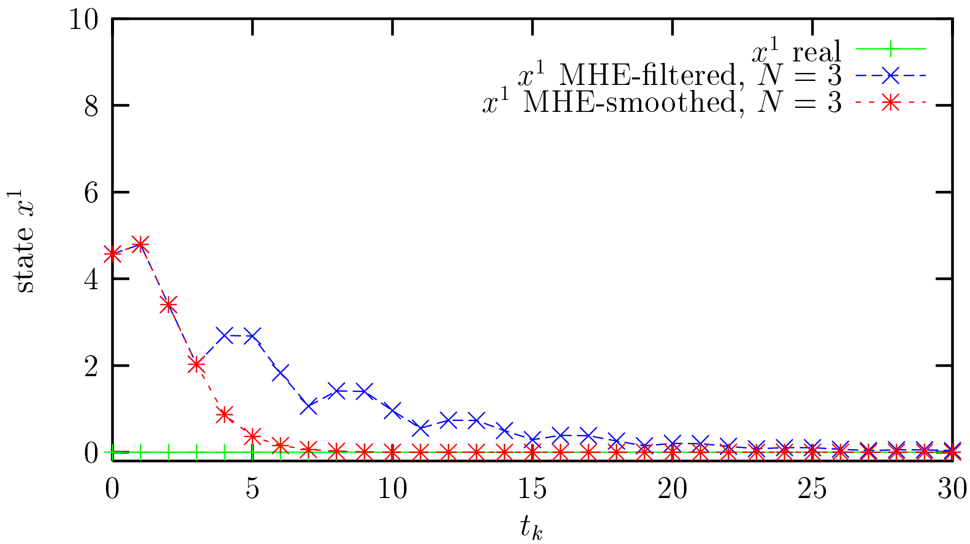

Figure 3.12: Example 2.: Stability in dependence of Q_{-N}^{-1}, $N = 3$.

Example 3., the Influence of N, Non-Intuitive Tuning

The last example was chosen, to clarify the fact, that the tuning of the MHE with filtered or smoothed update is often nonintuitive. To show this fact, the following modified system of second order was used

$$\begin{bmatrix} 0.1 & 1.1 \\ -0.95 & 0 \end{bmatrix} x_k$$

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k$$

$$Q^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad R^{-1} = 1/100$$

$$Q_{-N}^{-1} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

$$\text{First Picture} \quad N = 7$$

$$\text{Second Picture} \quad N = 12$$

We see that the eigenvalues $\lambda_{1,2} = 0.05 \pm 1.021i$ are conjugated complex and have an absolute value $|\lambda| = 1.0222 > 1$. The system shows oscillatory behavior. R^{-1}, Q_{-N}^{-1} are chosen as

$$Q^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad R^{-1} = 1/100 \quad Q_{-N}^{-1} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

For these values the maximum eigenvalue of Λ as given in Theorem 3.3 and Theorem 3.4 for different N's was calculated . The results are given in Table 3.3.5. Interesting is, that for the filtered MHE the tuning via N is non-intuitive. For example an horizon length of 2 leads to a stable estimator, whereas a horizon length of 3 leads to an unstable one. This continues until a "critical" horizon length of 7 is reached, hereafter all horizon length of N are stable (this is in the table only shown for a horizon length up to 10, however further calculations up to $N = 100$ confirmed these result). Figure 3.13 and Figure 3.14 show simulations for $N = 7$ (filtered MHE unstable) and Figure 3.14 show simulations for $N = 12$ (filtered MHE stable). The smoothed MHE is for all values of N stable for the given system and weights. The maximum of amplitude of the state x^2 in Figure 3.13 for the filtered MHE is strictly increasing, whereas the maximum amplitude for the smoothed MHE is decreasing. A possible explanation of this fact might be, that the filtered MHE can be seen as N parallel running estimators. If the horizon length of the MHE is chosen so, that it somehow corresponds, or coincides with the frequency of the autonomous system, then instability could occur.

N	filtered $\|\lambda\|_{max}$	smoothed $\|\lambda\|_{max}$
1	1.03470	0.96753
2	0.97068	0.94370
3	1.04680	0.91779
4	0.92740	0.89676
5	1.03500	0.87415
6	0.87010	0.85613
7	1.00060	0.83704
8	0.80309	0.82204
9	0.94730	0.80630
10	0.73070	0.79416

Table 3.2: Maximal eigenvalues in dependence of horizon length N

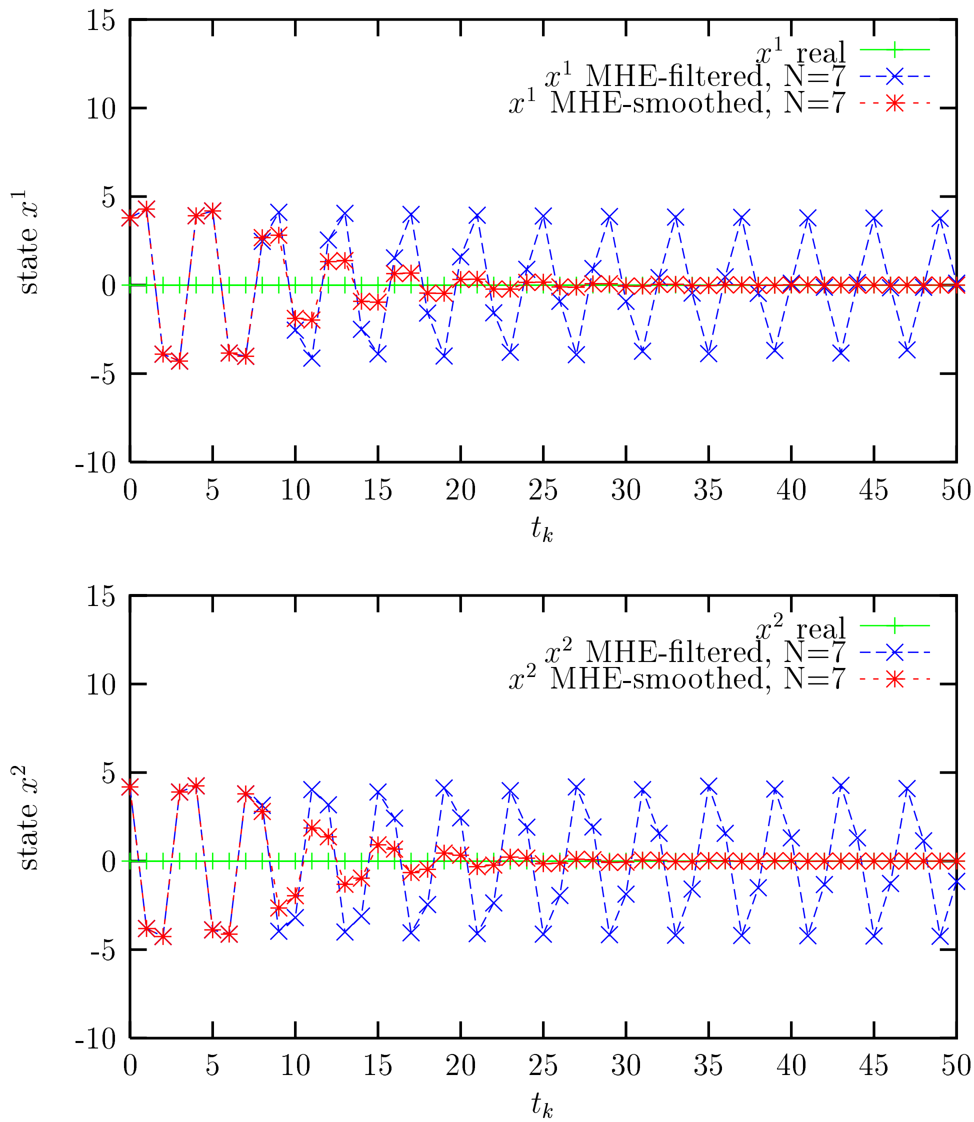

Figure 3.13: Example 3.: Stability in dependence of N, $N = 7$.

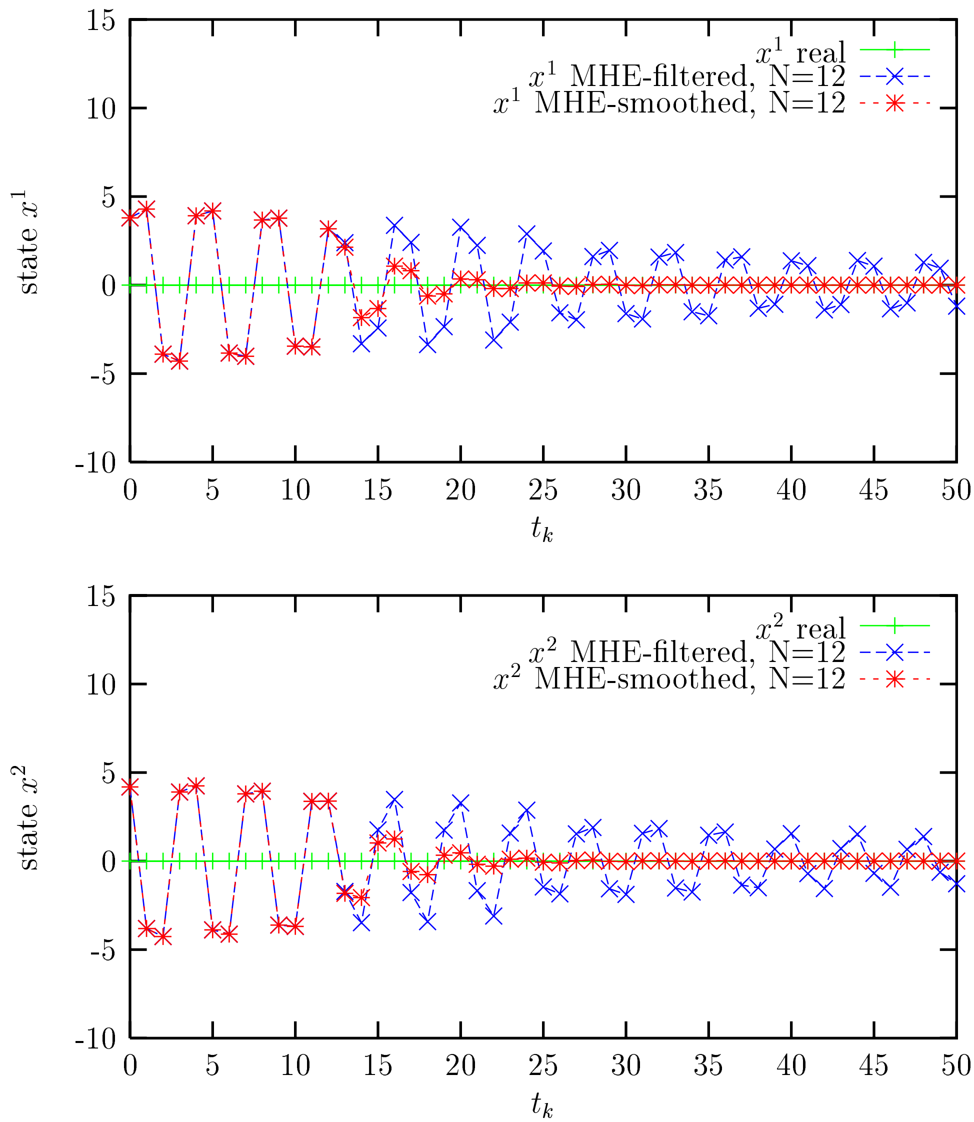

Figure 3.14: Example 3.: Stability in dependence of N, $N = 12$.

3.3.6 Summary of MHE stability results

This part tries to summarize "all" so far found stability results for the MHE with filtered and smoothed updates. It serves as a motivation for the next sections ideas on the use of Lyapunov function arguments in order to show stability.

Stability of MHE with filtered update So far the following stability conditions for MHE with filtered update are found:

- Choice of $Q_{-N}^{-1} = 0$ guarantees stability for all horizon lengths N. This result was presented in the review chapter 2.4.3 and guaranteed stability even if constraints were considered. However it was found that this method "trivializes" the MHE stability question by removing the connection to previous estimates.
- Stability if $Q_{-N} = \tilde{Q}_{ARE}$. This follows directly from the equivalence of the MHE with filtered update and the KF derived in section 3.1. The KF is stable if P is taken as the solution of the discrete algebraic Riccati equation (DARE) [6].
- For Q_{-N} satisfying the filtering stability condition given in Theorem 3.3 stability can be guaranteed. This choice has the disadvantage that estimator "cycling" shows up. Additional, and similar to MPC [7] strategies, no intuitive way to decide if for an longer horizon length N or slightly different Q_{-N} stability is preserved if stable initial values are known. This behavior was also examined in the previously given examples. It is often referred to as "non-intuitive" with respect to the tuning parameters.

Smoothed update schemes For Q_{-N} satisfying the smoothing stability condition given in Theorem 3.4 stability can be guaranteed. This choice has the advantage that no estimator "cycling" shows up. However the non-intuitive character of the tuning parameters, similar as for the filtered update, still exists. This behavior was examined also in the previously given examples, however no direct example showing this was given. An example showing this could probably be deduced from the MPC counterexample given by Bitmead*et al.* [7].

Time-variant Q_{-N} If time varying $Q_{-N|k}$ are considered stability can be guaranteed by taking as initial x-value the filtered prediction and as $Q_{-N|k}$ the covariance prediction of the KF at time k. This result follows directly from the in section 3.1 derived equivalence between the KF and MHE with KF-like update.

3.3.7 A Lyapunov Function for MHE

The stability conditions for the MHE given in the previous sections are not very satisfying. An extension to the constrained case is not directly possible. Therefore it seems to be important that other starting points for the derivation of the stability of the MHE should be investigated. Ideas which come to mind are the use of Lyapunov functions in order to guarantee stability like in the LQR or MPC formulations . In the MPC formulation a natural choice is the use of the value-function [5]. The use of the value-function gives an easy and descriptive way to understand how MPC strategies achieve stability and feasibility. The main argument used is that the value of the cost-function is decreasing from step to step. If the general MHE value-function is examined as a possible Lyapunov function, it can be shown that for all $Q_{-N}^{-1} \geq Q^{-1}$ a M exists, such that for $N \geq M$ the value function is valid Lyapunov function for MHE with smoothed update scheme. This will be examined in the following

$\Psi_k^\star$ as Lyapunov function for the smoothed update scheme

For the MPC problem a natural choice of a Lyapunov function is the value function $\Phi_k^\star$. However similar to the MHE problem, not every choice of N and $\tilde{Q}_N$ does imply that $\Phi_k^\star$ satisfies the "stability" condition $\Phi_k^\star - \Phi_{k+1}^\star < 0$[6]. For the MPC case there exist different ways to address this "decrease" in the value function. A typically way, to set $x_{k+N|k} = 0$ at the end of the horizon what additional results in $u_{k+N-1+j|k} = 0$, $i > 0$ [27]. This guarantees, that one non-optimal feasible solution at time $k+1$ is $\{u_{k+1|k}^\star, \cdots, u_{k+N-1|k}^\star, 0\}$. Use of this solution leads to the hypothetical non-optimal cost function $\tilde{\Phi}_{k+1} > \Phi_{k+1}^\star$. Since $\tilde{\Phi}_{k+1} = \Phi_k^\star - x_k^T \tilde{Q} x_k - {u_{k|k}^\star}^T \tilde{R} u_{k|k}^\star > \Phi_{k+1}^\star$ it follows that $\Phi_k^\star - \Phi_{k+1}^\star > 0$. The direct transfer of this approach to the MHE case is not possible, since this would require to set $e_{k|k} = 0$, which corresponds to a trivial solution. Another approach used for MPC is to consider a final state penalty Q_N [7][7]. However these approaches are also difficult to apply to the MHE problem, since they either lead to a trivial solution, or they give useless solutions.

[5]Note the important difference between the use of value function and cost-function. A value-function is the cost-function evaluated in dependence of an optimal x. The value function has no dependence on the time k, whereas the cost-function has.

[6]This is not the only property a Lyapunov function must have. From Theorem 1.1 we know, that additional $\Phi_k^\star$ must be bounded below by a $\mathcal{K}^+$ function $\alpha(\|x_k\|) \leq \Phi_k^\star(x_k)$, this is satisfied, since $x_k^T Q x_k \leq \Phi_k^\star$. The third necessary property is that $\Phi_k^\star$ must be upper bounded by a $\mathcal{K}^+$ around the origin. This is satisfied, since from the analytical solution of the unconstrained MPC problem we know, that $\Phi_k^\star = x_k^T \Pi_0 x_k$

[7]This can be also seen as setting $x_{k+N|k} = 0$ and the consideration of a infinite prediction in the future [34]

Another for the MHE problem more suited method bases on the following well known fact for the MPC problem: For the MPC formulation with $\tilde{R} > 0$, $\tilde{Q} \geq 0$, $(A, \tilde{Q}^{1/2})$ observable,(A, B) stabilizable there is for *every* $\tilde{Q}_N$ a N so that $\Phi_k^\star$ satisfies $\Phi_k^\star - \Phi_{k+1}^\star > 0$.

Outline of proof:

$\Phi_k^\star - \Phi_{k+1}^\star$ can be formulated as follows

$$\Phi^{N^\star}(x_k) \;-\; \Phi^{N^\star}(x_{k+1}) \;=\; x_k^T\tilde{Q}x_k \;+\; {u_{k|k}^\star}^T\tilde{R}u_{k|k}^\star \;+\; \underbrace{\Phi^{N-1^\star}(x_{k+1}) - \Phi^N(x_{k+1})}_{x_{k+1}^T\left(\Pi_0^{N-1}-\Pi_0^N\right)x_{k+1}}$$

Rewriting of $\Phi^{N^\star}(x_k)$ as

$$\Phi^{N^\star}(x_k) = x_k^T\tilde{Q}x_k + {u_{k|k}^\star}^T\tilde{R}u_{k|k}^\star + \Phi^{N-1^\star}(x_{k+1})$$

is possible, since the input sequence without the first element, which is optimal for N at k is also optimal at $k+1$ (Bellman Principle of Optimality). A standard result is [3], that for $N \to \infty$: $\Pi_0^N \to \Pi_\infty$, where Π_∞ is the solution of the algebraic Riccati equation, as given in 3.27. By this $\Pi_0^{N-1} - \Pi_0^N \to 0$ as $N \to \infty$. This implies, that there exist a M for every $\tilde{Q}_N$ such that $\Phi_k^\star - \Phi_{k+1}^\star > 0$ if $N \geq M$ which completes the proof. □

Direct application of Bellman's Principle of Optimality is not possible, since Q_{-N}^{-1} in the error equation (3.44) shows up at the beginning. This implies that the sequence $\{{w_{k-N|k}^N}^\star, {w_{k-N+1|k}^N}^\star, \ldots, {w_{k-1|k}^N}^\star\}$[8] resulting from $\Psi_k^{N^\star}(e_{k-N-1|k}^\star)$ is none optimal for $\Psi_k^{N-1^\star}(e_{k-N|k}^\star)$.

$$\Psi_k^{N^\star}(e_{k-N-1|k-1}^\star) \neq {w_{k-N-1|k}^\star}^T Q_{-N}^{-1} w_{k-N-1|k}^\star + {e_{k-N|k}^N}^{\star T} C^T R^{-1} C {e_{k-N|k}^N}^\star + \Psi_k^{N-1^\star}(e_{k-N|k}^\star)$$

However under the assumption that $Q_{-N}^{-1} \leq Q^{-1}$, $Q^{-1} > 0$, $C^T R^{-1} C \geq 0$, $(A, CR^{-1/2})$ observable[9] the following can be shown.

Theorem 3.5 *Suppose $Q_{-N}^{-1} \leq Q^{-1}$, $Q > 0$, $C^T RC \geq 0$, $(A, CR^{-1/2})$ observable, then there exists a M such that $\forall N \geq M$ $\Psi^{N^\star}(e_{k-N-1|k}^\star)$ is a Lyapunov function for $e_{k-N-1|k-1}^\star$ and $e_k = e_{k|k}$ is asymptotically stable.*

[8] $\hat{w}$ was replaced by w for simplicity reasons

[9] These assumptions are necessary, to guarantee convergence of the Riccati equation (3.49)-(3.50) for Υ. They are the dual counterparts to the in Theorem 3.2 given conditions, $\tilde{Q} \geq 0$, $\tilde{R} > 0$, $(A, \tilde{Q}^{1/2})$ for MPC. Compare also (3.44) and (3.25).

Proof: The proof runs similar to the one for the MPC case. First a upper bound for $\Psi^{N\star}(e^\star_{k-N-1|k-1})$ in Terms of ${w^N_{k-N-1|k-1}}^\star$, ${e^N_{k-N|k}}^\star$ and ${\Psi^{N-1}_k}^\star(e^\star_{k-N|k})$ is calculated.

$$
\begin{aligned}
&\Psi^{N\star}_k(e^\star_{k-N-1|k-1}) = \\
&{w^N_{k-N-1|k}}^{\star T} Q^{-1}_{-N} w^\star_{k-N-1|k} + {e^N_{k-N|k}}^{\star T} C^T R^{-1} C {e^N_{k-N|k}}^\star \\
&\underbrace{+{w^N_{k-N|k}}^{\star T} Q^{-1} {w^N_{k-N|k}}^\star - {w^N_{k-N|k}}^{\star T} Q^{-1}_{-N} {w^N_{k-N|k}}^\star}_{\geq 0 \quad \text{since} \quad Q^{-1}_{-N} \leq Q^{-1}} \\
&\underbrace{\begin{array}{l} +{w^N_{k-N|k}}^{\star T} Q^{-1}_{-N} {w^N_{k-N|k}}^\star \\ +\sum_{j=1}^{N-1}\left({w^N_{k-N+j|k}}^{\star T} Q^{-1} {w^N_{k-N+j|k}}^\star + {e^N_{k-N+j|k}}^{\star T} R^{-1} {e^N_{k-N+j|k}}^\star\right) + {e^N_{k|k}}^{\star T} R^{-1} {e^N_{k|k}}^\star \end{array}}_{\geq {\Psi^{N-1}_k}^\star(e^\star_{k-N|k}) \quad \text{since} \quad {w^N_{k-N+j|k}}^\star \quad \text{is not optimal for} \quad \Psi^{N-1}_k}
\end{aligned}
$$

from this it follows, that

$$
\Psi^{N\star}_k(e^\star_{k-N-1|k-1}) \geq w^\star_{k-N-1|k} Q^{-1}_{-N} w^\star_{k-N-1|k} + {e^N_{k-N|k}}^{\star T} C^T R^{-1} C {e^N_{k-N|k}}^\star + {\Psi^{N-1}_k}^\star(e^\star_{k-N|k})
$$

Now the same procedure as for the MPC case can be applied. Calculating the difference of two preceeding value functions gives:

$$
\begin{aligned}
\Psi^{N\star}_k(e^\star_{k-N-1|k-1}) - {\Psi^N_{k+1}}^\star(e^\star_{k-N|k}) \geq\ & w^\star_{k-N-1|k} Q^{-1}_{-N} w^\star_{k-N-1|k} + {e^N_{k-N|k}}^{\star T} C^T R^{-1} C {e^N_{k-N|k}}^\star \\
& \underbrace{+{\Psi^{N-1}_k}^\star(e^\star_{k-N|k}) - {\Psi^N_{k+1}}^\star(e^\star_{k-N|k})}_{{e^\star_{k-N|k}}^T(\Upsilon^{N-1}_{-N} - \Upsilon^N_{-N-1}) e^\star_{k-N|k}}
\end{aligned}
$$

Similar as for MPC $\Upsilon_{-N-1} \to \Upsilon_{infty}$ for $N \to \infty$. From this it follows, that $\exists$ M so that $\forall N \geq M$ $\Psi^{N\star}(e^\star_{k-N-1|k})$ is a Lyapunov function for $e^\star_{k-N-1|k-1}$[10]. This also implies, that $\Psi^{N\star}_k$ is a Lyapunov function for the error $e_k = e_{k|k}$ so that e_k is asymptotically stable. This completes the proof. □

[10] Since $\Psi^{N\star}_k(e^\star_{k-N-1|k-1}) - {\Psi^N_{k+1}}^\star(e^\star_{k-N|k}) \geq 0$ and from the analytical solution of Ψ^N_k it is known, that properties 1.) and 2.) of Theorem 1.1 are satisfied

Chapter 4

Conclusions

This work has focused on 3 goals: to provide an as complete and accurate review of MHE state estimation methods; to derive and clarify some important results for the linear MHE problem without the consideration of constraints; and to supply an easy to use software tool for the simulation of MHE algorithms.

The review section gives a self contained overview on existing MHE estimation methods for discrete time linear and nonlinear systems. One of the results of this presentation is that the feasibility of future estimates seems to be more difficult than for the MPC problem, even if only the nominal problem is considered. In contrast to the MPC approach, it does not make sense to apply a zero endpoint constraint, since this would mean setting the nominal error for the actual filtered estimate to 0. In the MPC problem this is possible since a prediction in the future is made, and by this a restriction on the states at the end of the horizon does not imply that the system goes to the origin at the next step. There seem to be only a few possibilities to overcome this problem. One might be to predict additionally in the future, as proposed by [49, 31]. The most important open question is how to overcome the stability problems for estimation schemes that do not consider a Kalman filter like update. One result of this lack of a general stability scheme is that there exist only two nonlinear MHE/BE schemes that can guarantee stability, the free floating MHE as proposed by [30, 35] and the nonlinear Batch state estimator. This problem was the main motivation for the examination of the stability of the finite horizon linear MHE given in chapter 3.

During the theory part, the connections between the KF, MHE, and the batch filter were derived. This understanding is important, because most stability proofs for the batch estimator and the MHE (with specific update schemes) are based on the KF. In order to clarify the connections, an analytic solution for the MHE problem was derived so that the resulting equations can be reformulated in KF–like form. The KF and the MHE with KF–like update can be seen as "recursive" solution methods for the batch filter. The preceding part clarified the connections between the MPC problem and the MHE problem. It was shown that both problems are nearly "dual". As a matter of facts for the infinite horizon case the MHE can be seen as the dual of the MPC. For the finite horizon this is not possible, since the given setup for the MHE contains and/or misses terms which are inherent for the MPC problem. This leads to slight differences in the resulting Riccati equations for the calculation of the gain matrices. Based on these analytic solutions for the filtering matrix, new stability conditions for the linear, unconstrained finite horizon MHE with filtered and smoothed updates were derived. Examples showed that the tuning of the resulting estimators is non-intuitive. For example, it is possible that an MHE is stable for a specific horizon length, but unstable for a horizon length that is bigger than the stable one. This behavior can also be seen in the case of finite horizon MPC without a terminal constraint or additional "infinite" cost/penalty. Since the presented method for proving stability allows no direct transfer to the nonlinear or constrained case, some preliminary investigations concerning the form of a cost function for MHE as a Lyapunov function were presented in order to overcome this problem.

The MHE toolbox provided allows the easy simulation of different MHE concepts. The toolbox was designed to offer easy integration and combination with other programs for simulation and calculation of plant models and controllers. One such integration would be the coupling of the MHE toolbox with the NMPC toolbox, which would allow the simulation of a (nonlinear) moving horizon controller/estimator combination. Such a tool was not available previously, and should allow easier investigation of MHE algorithm properties. All of the given simulations were calculated with the help of this toolbox and are included as explanatory examples.

This work raises a number of theoretical and practical issues as starting points for further investigations.

- Section 3.3.5 provides some preliminary thoughts about the cost function as a Lyapunov function. However, this seems to be one of the unresolved key steps for a unified stability proof for MHE schemes for nonlinear and/or constrained systems. Similar to MPC, the cost function as Lyapunov function would offer the possibility to derive new algorithms. This would also allow an extension to more general (nonlinear) cost functions. One of the key steps for the cost function being a Lyapunov function in the MPC concept is to consider an infinite prediction horizon or a final, zero endpoint constraint. Both ways seem to be not promising for the MHE problem, since the MHE problem considers information in the past, whereas for the MPC problem a prediction in the future is made. This prediction in the MPC scheme allows one to set the state to 0 at the end of the horizon without "instantaneously influencing" the actual state, whereas a similar idea would lead to a trivial solution, for the MHE problem, by setting the nominal error to zero.

- As shown in chapter 2, the question of feasibility of future estimates, even for the nominal case, is still mostly unanswered. For a rigorous stability proof for constrained MHE, however, the understanding and solution of this obstacle is necessary. For the MPC problem, an infinite horizon or a zero endpoint constraint can provide a solution, since this guarantees that a new feasible solution in the future exists. For the MHE this is problematic, since no predictions are made in the future. There are only a few algorithms [49, 31], which consider additional to the estimation horizon in the past (over the measurement values) a prediction in the future, and try to overcome by this the feasibility problem in a similar way as in MPC methods. Other solutions might be possible.

- An MHE scheme with a more general cost function as proposed in section 2.3.3 could provide the possibility to consider “soft–constraints” similar to nonlinear MPC [11] on estimated states and inputs. Unrealistic states could be weighted much more heavily (not only in a quadratic sense) without applying hard constraints. From a probabilistic point of view, this might offer the possibility to consider non normal distributions.

- The integration of MHE and MPC in a closed algorithm might allow improved behavior and easier stability proof. This is especially important, since no separation principle exist for the nonlinear estimation/control problem. The performance might improve, especially if sensitivity information of the MPC could be used to decide which state needs to be estimated “best.” For example in the case of a only stabilizable system (e.g. a system with uncontrollable but stable modes) it might be possible, that not the whole state vector must be estimated for control purposes. Another way is to look at the “separated” combination of MPC and MHE as employed by Scokaert et. al. [43]. They consider the discrete time stability problem under perturbations and derive under additional assumptions that the coupling of MPC and state estimation can lead to a stable system. This can be seen as a generalization of the separation principle. Ways to derive and propose an stable, real integrate MPC/MHE combination might be possible.

- The supplied tool box allows room for further improvements. The consideration of general nonlinear cost functions and nonlinear constraint as given in section 2.3.3 could be considered for future work.

Appendix A

General Theorems and Proofs

This part contains some general theorems and proofs which would hinder the natural flow of the arguments.

A.1 The Probability Transfer Function Theorem or derived density

Theorem A.1 (Probability transfer function theorem) *Let x, $y \in \mathbb{R}^n$ be random n-vectors. Suppose that $y = f(x)$ and f^{-1} exists and that f and f^{-1} are continuous differentiable, $f,\ f^{-1} \in C^1$. Then*

$$p_y(y) = p_x\left(f^-1(y)\right)\left\|\frac{\partial f^{-1}(y)}{\partial y}\right\|,$$

with $\left\|\frac{\partial f^{-1}(y)}{\partial y}\right\|$ being the absolute value of the Jacobian determinant.

Proof: See [16] Theorem 2.7. □

Appendix B

NMHE Toolbox

The nonlinear moving horizon state estimation toolbox was developed as tool to investigate the properties of MHE schemes. All of the examples given in this work are calculated using this toolbox. The code is written in Octave [10] which is a high level MATLAB like interactive language for numerical computations. To run the toolbox, a octave version greater than 2.0.5 is required. The code was developed with the goal to allow an easy incorporation into closed loop simulations. By this allowing the use in already established programs for closed loop controller simulations.

B.1 Considered Estimators and Systems

The code is intended for state estimation of the following class of noisy, time-variant, discrete-time nonlinear systems:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k, k) + w_k \qquad k = 0, 1, 2, \ldots \\ y_k &= g(x_k, k) + v_k \end{aligned}$$

$$\begin{aligned} x_k &\in \mathbb{R}^n && \text{states of the system} \\ y_k &\in \mathbb{R}^p && \text{measured output} \\ w_k &\in \mathbb{R}^n && \text{state noise} \\ v_k &\in \mathbb{R}^p && \text{measurement noise} \end{aligned}$$

In order to allow the simulation of a wide class of NMHE formulations using quadratic cost functions, a broad approach was used. MHE formulations that can be expressed in terms of equations (B.2)- (B.4) can be simulated. The following properties are noteworthy:

- The initial state is the one step ahead prediction of the filtered or smoothed state $\hat{x}^{\star}_{k-N-1}$ estimate

- The initial weight $Q_{-N|k}$ is calculated from the last initial weight $Q_{-N|k-1}$, the weighting matrixes R, Q, Q_0 the actual discrete valued time k, and the horizon length N using a user supplied function. If no such function is given the initial weight Q_0 is taken.
- It is possible to incorporate nonlinear constraints on the states x_k, linear constraints on the state errors w_k and the measurement errors v_k.

With this the used MHE formulation can be given as follows:

$$\begin{aligned}\min_{\{\hat{w}_{k-N-1|k},\cdots,\hat{w}_{k-1|k}\}} \Psi_k : \quad \Psi_k &= \hat{w}^T_{k-N-1|k} Q^{-1}_{-N|k} \hat{w}_{k-N-1|k} \\ &+ \sum_{j=k-N}^{k-1} \hat{w}^T_{j|k} Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}^T_{j|k} R^{-1} \hat{v}_{j|k}\end{aligned} \tag{B.1}$$

subject to the state equality constraints:

$$\begin{aligned}\hat{x}_{k-N|k} &= \bar{x}_{k-N|k} + \hat{w}_{k-N-1|k} \\ \text{with} &\quad \bar{x}_{k-N|k} = f(\hat{x}^\star_{k-N-1|k-N-1}, u_{k-n-1}, k) \quad \text{filtered prediction} \\ \text{or} &\quad \bar{x}_{k-N|k} = f(\hat{x}^\star_{k-N-1|k-1}, u_{k-n-1}, k) \quad \text{smoothed prediction} \\ \hat{x}_{j+1|k} &= f(\hat{x}_{j|k}, u_j) + \hat{w}_{j|k} \quad , j = k-N \dots k-1 \\ y_j &= g(\hat{x}_{j|k}, k) + \hat{v}_{j|k} \quad , j = k-N \dots k\end{aligned} \tag{B.2}$$

The following inequality constraints can be enforced:

$$\begin{array}{ccccccc} x_{nllo} & < & F(x_j) & < & x_{nlup} & j = & k-N \quad ,\dots,k \\ w_{lo} & < & \tilde{A}w_j & < & w_{up} & j = & k-N-1,\dots,k-1 \\ v_{lo} & < & \tilde{B}v_j & < & v_{up} & j = & k-N \quad ,\dots,k \end{array} \tag{B.3}$$

$Q_{-N|k}$ is calculated via a user supplied function:

$$Q_{-N|k} = q(Q_{-N|k-1}, Q, R, A, C, N) \tag{B.4}$$

B.2 General Overview/Features

The nmhe-toolbox routines are designed for easy incorporation in closed loop simulations. This means that once the necessary startup informations is entered by calling *nmhe_init* (this function sets the basic parameters) and *nmhe_weights* (this sets the weights for the quadratic objective) new estimate can be calculated. This state estimation calculation is done in a repeated manner by calling *nmhe_main* (main state

estimation routine) with new available measurements (which are possible results of simulating a plant model) and the actual control that is applied to the system. Necessary old measurement values are "stored" in the nmhe-routine internally in the octave structure variable *nmhe*; therefore this variable name is not available for the user. An overview-diagram on the use of the nmhe-toolbox for closed loop simulations is given in figure B.1. The first call to the toolbox is the *nmhe_init* routine which allows to set basic informations like the initial estimate for the state, the horizon length and function names for measurement calculation and state propagation. Then the weights for the cost function, the update scheme for $Q_{-N|k}$ and the choice between a smoothed or filtered update for $\hat{x}^{\star}_{k-N-1}$ is made by calling *nmhe_weights*. Constraints on states, state errors and measurement errors can be included using the optional *nmhe_?constraints* functions. Options concerning the amount of information that is printed out during the solution process and also the choice if data should be saved into a file can be made using *nmhe_options*. After this the actual estimates are calculated by repeatedly calling the *nmhe_main* routine. This routine takes the actual measurements and the actual applied control variable value and adds these to an internal "storage". Using now the stored values the nonlinear program given by equations (B.2)- (B.4) is build up and handed to an nonlinear optimization routine in order to calculate the new estimate. The nonlinear optimizer sued is the popular NPSOL algorithm [14].
Important features of the nmhe-toolbox are.

- The user can supply a function for the calculation off $Q_{-N|k}$ in dependence of old values.
- Use of filtered or smoothed estimate for initial estimate calculation.
- Possibility to use continuous time models for the calculation of new state estimates.
- Possibility to save "all" intermediate variables automatically into a file for later investigation.

B.3 Description of Routines

The routines are divided into three classes; criteria for division is the handling requirement by the user. All routines that must be called at minimum once by the user are in the "necessary" class. Routines that can be called by the user if specific settings are required are in the optional class and routines that are not called directly by the user are in the hidden routines class. The following descriptions are only intended

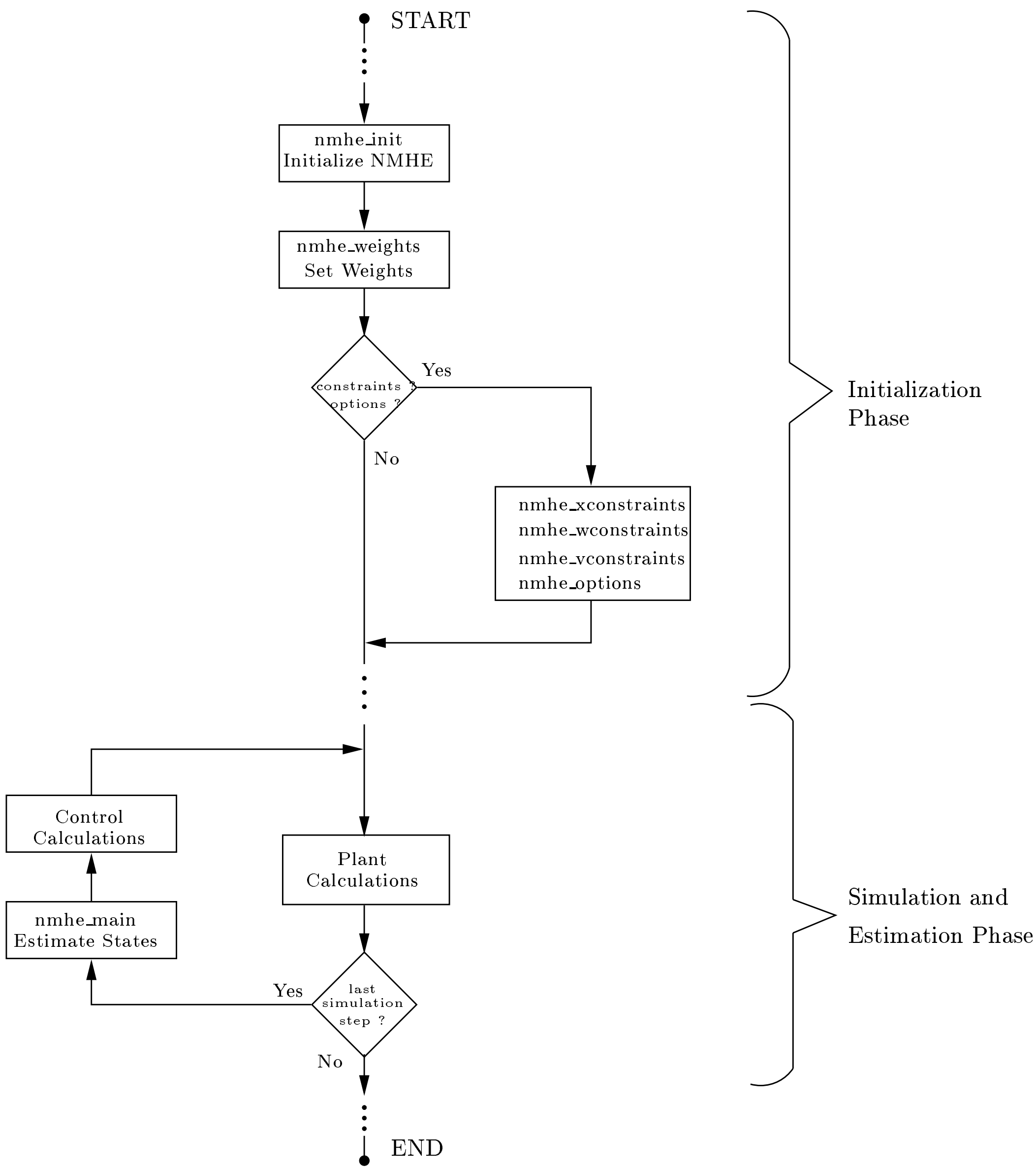

Figure B.1: NMHE toolbox, use in closed loop simulations.

to give a basic overview on the nmeh-toolbox. If more documentation is needed the different subroutines should be consulted, since most of the code is well documented. Note: Variables in [] are variables/values which are optional. If more then one optional variable is used then it is important to to keep the order of the variables in the function calls.

Necessary Routines

Necessary Routine	Description
nmhe_init.m	set and initialize the basic values
nmhe_weights.m	set the weights needed for the cost-function calculation
nmhe_main.m	main estimation routine

Table B.1: Necessary Routines

nmhe_init.m This is the main initialization routine for the toolbox. All necessary variables are set up and initial values are assigned. This routine must be called first. All old values of nmhe-intern variables used in previous runs are cleared when this function is processed. The header of the file is given in Header B.3.

X0BAR is the initial guess $\bar{x}_0$, *UO* is the control step taken one time-step before the estimator is started (0 for most cases), *N* is the horizon length of the estimator. *DELT* is the sampling time of the system, this value is only important if the next state is calculated by integrating an continuous-time state model. *STARTTIME* is the actual time at the start of the estimation process. This value is of interest if continuous time systems or time variant systems are considered. *FNAME* is a string specifying the octave file in which the model for the state equations of the examined systems are given. The use of an external function to calculate the next state allows more flexibility. It is possible to calculate the propagated state by integrating a continuous-time model if an analytical discrete time model is not available. The following variables are handed to the function given in *FNMHE* in the order presented: the old system state *XOLD*, the piecewise constant input *U*, the time *TIME* and the sample-time interval *SAMPELTIME*. *GNAME* is a string specifying the octave file which contains the model measurement equations. For this calculation the state vector *X* and the time *TIME* at which the measurement should be calculated are passed through. Examples for *FNMHE* propagation functions and *GNMHE* measurement functions are given in Program B.3.

Header B.1 Header of nmhe_init.m

```
## usage: [] = nmhe_init (XOBAR, U0, N, DELT, STARTTIME, FNAME, GNAME
##                                                                [, XINIT])
##
## Initialization function for the nonlinear moving horizon estimator
##
## XOBAR     = initial x0 estimate
## U0        = value of u one step before t=0, or if no value u is
##             given using nmhe_main, steady feed value
## N         = horizon length
## DELT      = sample time length
## STARTTIME = start time for estimation
## FNAME     = system function name
## GNAME     = output function name
## XINIT     = binary variable if set to 1 the smoothed update
##             for x_k-N-1|k is used, default values is 0
##             corresponding to an filtered update
##
```

Program B.1 Sample files for the state propagation and measurement functions

```
function XNEW=fsystem(XOLD,U,TIME,SAMPLETIME)
 .
 .
 ## calculation of propagated state
 .
 .
 XNEW= ..  ## set return value

endfunction

function Y=gsystem(X,TIME)
 .
 .
 ## calculation of new y's
 .
 .
 Y= ..  ## set return value

endfunction
```

nmhe_weights.m With this routine the weights needed for the calculation of the quadratic cost-function given in B.2 can be set. The routine header is given in Header B.3. *QN0INV* specifies the initial weight Q_0^{-1} on $\hat{w}_{0|k}$, $k = 0 \dots N$ needed

Header B.2 Header of nmhe_weights.m

```
## usage: [] = nmhe_weights (QN0INV, QINV, RINV [, QNFNAME])
##
## Initializes the weight matrixes for the nonlinear moving horizon
## estimator. w's are state disturbances, v's are output-disturbances
## QNFNAME specifies which function to use for the Q_-N|k update
##
## QN0INV   = Initial startup weight matrix for w0
## QKINV    = general weight matrix for wj|k
## RINV     = general weight matrix for vj|k
## QNFNAME  = optional string, specifies which function to use
##            for Q_-N|k update calculation,  if omitted QN0INV
##            is used
```

during the batch phase of the NMHE. This value is also used as an initial weight $Q_{-N|k}^{-1}$ during the moving phase if no function for the calculation of this matrix is supplied. The function name for the calculation of $Q_{-N|k}$ can be passed to the toolbox routines using the optional variable *QNFNAME*. This user supplied function calculates $Q_{-N|k}$ using the values for the last $Q_{-N|k-1}$, R^{-1}, Q^{-1}, the actual discrete time k and the horizon length N. An example file for the calculation of $Q_{-N|k}$ is given in Program B.3. The matrix *Qinv* is the weight Q^{-1} and *RINV* is the weighting matrix R^{-1}.

Program B.2 Example Routine for the Calculation of $Q_{-N|k}$

```
function QNnew=(QNOLD,Q,R,N,K)
   .
   .
  ## calculation of QNnew using the supplied values
   .
  QNnew= ..  ## set return value

endfunction
```

nmhe_main.m This function is the heart of the NMHE algorithm. The user calls this routine subsequently with new measurements $Y = y_k$ and the actual applied control move $U = u_k$. The routine then calculates the new optimal filtered estimate $XHATSTAR = \hat{x}^\star_{k|k}$, the connected measurement error $VHATSTAR = \hat{v}^\star_{k|k}$, the state error $WHATSTAR = \hat{w}^\star_{k|k}$ and the value of the cost-function $PHIHATSTAR = \Psi^\star_k$ under the optimal $\hat{x}^\star_{j|k}$ sequence *XHATSTARSEQUENCE*.

Header B.3 Header of nmhe_main.m

```
## usage: [XHATSTAR, VHATSTAR, WHATSTAR, PHIHATSTAR, XHATSTARSEQUENC]
##                                                     = nmhe_main (Y, U)
##
## Main function for the nonlinear moving horizon estimator
## initialize first with nmhe_init, nmhe_options and nmhe_xconstraints
##  nmhe_wconstraints, nmhe_vconstraints if necessary
##
## Y = new measurement vector
## U = input applied to system at last step
##
## XHATSTAR         = calculated new state estimate
## WHATSTAR         = w disturbance for calculated state
## VHATSTAR         = v disturbance for calculated state
## PHIHATSTAR       = value of "least squares objective" for
##                    optimal xhatstar
## XHATSTARSEQUENCE = calculated optimal x sequence
```

Additional/Optional Routines

Optional routines can be called by the user if necessary. Table B.3 gives an overview about functions belonging to this class.

Optional Routine	Description
nmhe_options.m	handles output options
nmhe_xconstraints.m	defines the x constraints
nmhe_wconstraints.m	defines the w constraints
nmhe_vconstraints.m	defines the v constraints

Table B.2: Optional Routines

nmhe_options.m This function allows the user to specify if and which sort of output should be displayed on the screen or saved to a file during the solution of the NMHE problem. The header of this function is given in Header B.3.
The value of *DATASAVE* indicates if information should be saved to the *DATASAVE-NAME* file.
The *INFO* variable sets the screen output level of the nmhe_main routine. The default value is 0 corresponding to the fact that no information is printed on the screen. A value of 1 means that at each call of nmhe_main the actual time k and a description of the estimator status (startup, batch, moving phase) is printed. A value of 2 additionally displays the values for the horizon length N, the new estimated state sequence X and the newly calculated filtered estimate. A value of 3 corresponds to a printout of all internal used values. This allows if necessary simple debugging of the code .
The variable *INFOOUTPUTSTREAM* sets the standard output stream of octave to the specified value. If more information on this feature is needed the octave manual should be consulted.

nmhe_xconstraints.m With this function optional constraints on the state variables x_k can be defined. To enable a faster solution of the estimation problem, a division into three different types of x constraints is made. This division allows to exploit the feature that the NPSOL code can solve partly linear constrained nonlinear programs faster than if an fully nonlinear formulation is used. The state constraints given in equations B.3 are divided in the following three classes:

- Simple lower and upper bounds on x: $x_{lo} < \hat{x}_k < x_{up}$
- Linear constraints on x: $Ax_{lo} < A_x\hat{x}_k < Ax_{up}$
- Nonlinear constraints on x: $x_{nllo} < F(\hat{x}_k) < x_{nlup}$

For information about the use of this function, see the header given in Header B.3.
The variable *XNLFNAME* gives the filename with which the nonlinear constraints can be calculated using the passed trough $\hat{x}_k$ value. The structure of this user supplied function is similar to the structure of the measurement function *GFNAME* needed for nmhe_init.m .

nmhe_wconstraints.m With this routine constraints on the estimated state errors $\hat{w}_k$ can be defined. These constraints are only linear, since a more general definition would increase the necessary program code and the time needed for the solution of the

Header B.4 Header of nmhe_options.m

```
## usage: [] = nmhe_options (DATASAVE, DATASAVENAME, INFO
##                                        [,INFOOUTPUTSTREAM])
##
## Output options for the nonlinear moving horizon estimator
##
## DATASAVE         = binary value indicating if nmhe_main should keep
##                    track of all calculate values or only of for
##                    further calculation necessary values.
##                     0 == don't keep track of all values, save these
##                          values to file if an filename is given in
##                          DATASAVENAME, default setting
##                     1 == keep track and save all values in file at
##                          end of nmhe_main.
## DATASAVENAME     = filename in which to save values if datasave = 1
## INFO             = information level/printout level which
##                    nmhe_main is bringing up
##                     0 == no printouts, normal setting
##                     1 == basic printouts, e.g. k,..
##                     2 == 1 + printouts for x,Xi,N ...
##                     3 == 2 + printouts for internal variable nmhe
## INFOOUTPUTSTREAM = outputstream to use, normal setting = stdout
##
```

Header B.5 Header of nmhe_xconstraints.m

```
## usage: [] = nmhe_xconstraints ([XLO,XUP] [AXLO,AXUP,A]
##                                        [XNLLO,XNLUP,NLFNAME])
##
## Initializes the x constraints for the nonlinear moving horizon
## estimator different constraints types can be left out, but order
## of constraints must be maintained. If you don't have a lower or
## upper bound pick a large negative/positive number
##
## XLO = lower bound on x
## XUP = upper bound on x
##
## AXLO = linear lower bound on x
## AXUP = linear upper bound on x
## A    = linear constraints matrix
##
## XNLLO    = nonlinear lower bound on x
## XNLUP    = nonlinear upper bound on x
## XNLFNAME = nonlinear constraint function (string)
##
```

nonlinear program to much. The reason for this can be seen in the chosen implementation of the nonlinear program. The algorithm used optimizes over the states $\hat{x}_k$ and not over the $\hat{w}_k$ as in the problem statement. This change of optimization variables is possible since no G matrix in front of the w_k is considered. For the meaning of the different variables see the routine header given in Header B.3.

Header B.6 Header of nmhe_wconstraints.m

```
## usage: [] = nmhe_wconstraints (wlo, wup)
##
## Initializes the w (state errors) constraints for the nonlinear
## moving horizon estimator. If you don't have a value for a lower
## or upper bound pick a large negative or positive number instead.
##
## WLO = lower bound on w
## WUP = upper bound on w
##
```

nmhe_vconstraints.m This routine allows the definition of linear constraints on the estimated measurement errors $\hat{v}_k$. These constraints are also only linear. For the meaning of the different variables see the header of the routine given in Header B.3.

Header B.7 Header of nmhe_vconstraints.m

```
## usage: [] = nmhe_vconstraints (VNLLO, VNLUP)
##
## Initializes the constraints on the v (output errors) for the
## nonlinear moving horizon estimator. If you don't have a value for
## an lower or upper bound pick a large negative/positive number.
##
## VNLLO = lower bound on v
## VNLUP = upper bound on v
##
```

Hidden Routines

The hidden routines are necessary routines which make the handling of the toolbox by the user easier. The user is not confronted with these routines; these routines are only called by nmhe_main.m. The files are listed in Table B.3 and a short description follows.

Hidden Routine Description	
nmhe_phistart.m	calculates the cost-function value for startup-phase
nmhe_phibatch.m	calculates the cost-function value for batch-phase
nmhe_phi.m	calculates the cost-function value for moving-phase
nmhe_nlconststart.m	builds constraints for startup-phase
nmhe_nlconstbatch.m	builds constraints for batch-phase
nmhe_nlconst.m	builds constraints for moving-phase
nmhe_save.m	saves values as defined by nmhe_opts.m in file

Table B.3: Hidden Routines

nmhe_phi??.m This group including the nmhe_phistart.m, nmhe_phibatch.m, nmhe_phi.m functions is used to calculate the value of the costfunction for the corresponding phase. During these calculations the user supplied functions for the state and measurement model are used. A split-up into 3 different function was made to reduce the amount of IF constructs which must be evaluated during the nonlinear optimization.
These functions are called several times by the NPSOL optimizer in order to solve the nonlinear program.

nmhe_nlconst??.m Functions included in these group are nmhe_nlconststart.m, nmhe_nlconstbatch.m and nmhe_nlconst.m . These functions build and calculate the value of the constraints during the optimization process.

nmhe_save.m This routine saves the variables specified by the *DATASAVE* variable in nmhe_options.m in the file given by *DATASAVENAME*.

Known Problems

Known difficulties with the nmhe-toolbox are:

- The NPSOL optimization routine fails for ill-conditioned or “highly” nonlinear systems.

- Even for linear systems with long horizons or many states, the solution of one step during the moving phase can take up to several hours of computation time. For linear systems, a linear moving horizon toolbox is already written, but needs to be improved.

- Octave is still in a rapidly developing state, so that it could be necessary to rewrite some parts of the code if a version later than 2.0.7 is used.

B.4 Example Files

Several example files showing the use of the nmhe-toolbox are included in the subdirectory *examples* of the nmhe-toolbox path. Examples available include most of the simulations given in this work and additional examples showing the use of continuous time models.

Bibliography

[1] Richard E. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, New Jersey, 1957.

[2] B. Wayne. Bequette. Nonlinear predictive control using multi-rate sampling. *The Canadian Journal of Chemical Engineering*, 69, 1991.

[3] Dimitri P. Bertsekas. *Dynamic Programming.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[4] D. Bestle and M. Zeitz. Canonical form observer design for non-linear time-variable systems. *International Journal of Control*, 38(2):419–431, 1983.

[5] J. Birk and M. Zeitz. Extended Luenberger observer for non–linear multivariable systems. *International Journal of Control*, 47(6):1823–1836, 1988.

[6] Robert R. Bitmead and Michel Gevers. Riccati difference and differential equations: Convergence, monotonicity and stability. In Sergio Bittanti, Alan J. Laub, and Jan C. Willems, editors, *The Riccati Equation*, chapter 10, pages 263–291. Springer-Verlag, 1991.

[7] Robert R. Bitmead, Michel Gevers, and Vincent Wertz. *Adaptive Optimal Control, The Thinking Man's GPC.* Prentice-Hall, Englewood Cliffs, New Jersey, 1990.

[8] A. E. Bryson and Y. Ho. *Applied Optimal Control.* Hemisphere Publishing, New York, 1975.

[9] Carlos E. de Souza, Michel R. Gevers, and Graham C. Goodwin. Riccati equation in optimal filtering of nonstabilizable systems having singular state transition matrices. *IEEE Transactions on Automatic Control*, 31(9):831–838, September 1986.

[10] John W. Eaton. Octave—a high-level interactive language for numerical computations. Technical Report CCSR-93-003, The University of Texas Center for Control and Systems Research, Austin, Texas 78712, 1993.

[11] Rolf Findeisen. Suboptimal nonlinear model predictive control. Master's thesis, The University of Wisconsin–Madison, 1997.

[12] B. Friedland and I. Bernstein. Estimation of the state of a nonlinear process in the presence of nongaussian noise and disturbances. *Journal of the Franklin Institut*, 281:455–480, 1966.

[13] Arthur Gelb, editor. *Applied Optimal Estimation*. The M.I.T. Press, Cambridge, Massachusetts, 1974.

[14] Philip E. Gill, Walter Murray, Michael A. Saunders, and Margaret H. Wright. User's guide for SOL/NPSOL (Version 4.0): A Fortran package for nonlinear programming, technical report SOL 86-2. Technical report, Systems Optimization Laboratory, Department of Operations Research, Stanford University, 1986.

[15] Shi-Shang Jang, Babu Joseph, and Hiro Mukai. Comparison of two approaches to on-line prarameter and state estimation of nonlinear systems. *Industrial and Engineering Chemistry Process Design and Development*, 25:809–814, 1986.

[16] Andrew H. Jazwinski. *Stochastic Proceses and Filtering Theory*. Academic Press, New York, 1970.

[17] R. E. Kalman. Contributions to the theory of optimal control. *Bull. Soc. Math. Mex.*, 5:102–119, 1960.

[18] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Engineering*, pages 35–45, March 1960.

[19] R. E. Kalman and J. E. Bertram. Control system analysis and design via the "Second method" of Lyapunov, Part II: Discrete–time systems. *ASME Journal of Basic Engineering.*, pages 394–400, June 1960.

[20] R. E. Kalman and J. E. Bertram. Control system analysis and design via the "Second method" of Lyapunov, Part I: Continuous–time systems. *ASME Journal of Basic Engineering.*, pages 371–393, June 1960.

[21] Sathiya S. Keerthi. *Optimal Feedback Control of Discrete-Time Systems with State-Control Constraints and General Cost Functions*. PhD thesis, University of Michigan, 1986.

[22] I.W. Kim, M.J. Liebman, and T.F. Edgar. A sequential error-in-variables method for nonlinear dynamic systems. *Computers & Chemical Engineering*, 15(9):663–670, 1991.

[23] Arthur J. Krener and Alberto Isidori. Linearization by ouput injection and nonlinear observers. *Systems & Control Letters*, 3(1):47–51, 1983.

[24] Arthur J. Krener and Witold Respondek. Nonlinear observers with linearizable error dynamics. *SIAM Journal on Control and Optimization*, 23(2):197–216, 1985.

[25] H.J. Kushner. On the differential equations satisfied by conditional densities of markov processes. *SIAM Journal on Control and Optimization*, 2:106, 1964.

[26] Huibert Kwakernaak and Raphael Sivan. *Linear Optimal Control Systems*. John Wiley and Sons, New York, 1972.

[27] W. H. Kwon and A. E. Pearson. On feedback stabilization of time-varying discrete linear systems. *IEEE Transactions on Automatic Control*, 23(3):479–481, June 1978.

[28] Frank L. Lewis. *Optimal Estimation*. John Wiley and Sons, New York, 1986.

[29] David G. Luenberger. Observing the state of a linear system. *IEEE Trans. Mil. Elec.*, MIL-8:74–80, April 1964.

[30] Hannah Michalska and David Q. Mayne. Moving horizon observers and observer-based control. *IEEE Transactions on Automatic Control*, 40(6):995–1006, 1995.

[31] Kenneth R. Muske. *Linear Model Predictive Control of Chemical Processes*. PhD thesis, The University of Texas at Austin, 1995.

[32] Kenneth R. Muske and Thomas F. Edgar. Nonlinear state estimation. In Michael A. Henson and Dale E. Seborg, editors, *Nonlinear Process Control*, pages 311–370. Prentice Hall, 1996.

[33] Kenneth R. Muske and James B. Rawlings. Linear model predictive control of unstable processes. *Journal of Process Control*, 3(2):85–96, 1993.

[34] Kenneth R. Muske and James B. Rawlings. Model predictive control with linear models. *AIChE Journal*, 39(2):262–287, 1993.

[35] Kenneth R. Muske and James B. Rawlings. Nonlinear moving horizon state estimation. In Ridvan Berber, editor, *Methods of Model Based Process Control*, pages 349–365. Kluwer, 1995.

[36] Kenneth R. Muske and James B. Rawlings. Offset free model predictive control. 1997.

[37] Kenneth R. Muske, James B. Rawlings, and Jay H. Lee. Receding horizon recursive state estimation. In *Proceedings of the 1993 American Control Conference*, pages 900–904, June 1993.

[38] C. V. Rao, 1997. Personal Communication.

[39] James B. Rawlings. Advanced process dynamics and control. Course notes, University of Wisconsin, Madison, Spring 1996.

[40] James B. Rawlings and Kenneth R. Muske. Stability of constrained receding horizon control. *IEEE Transactions on Automatic Control*, 38(10):1512–1516, October 1993.

[41] Douglas G. Robertson, Jay H. Lee, and James B. Rawlings. A moving horizon-based approach for least-squares state estimation. *AIChE Journal*, 42(8):2209–2224, August 1996.

[42] Pierre O.M Scokaert and James B. Rawlings. Infinite horizon linear quadratic control with constraints. In *Proceedings of the 13th IFAC World Congress*, San Francisco, California, July 1996.

[43] Pierre O.M. Scokaert, James B. Rawlings, and Edward S. Meadows. Discrete-time stability with perturbations: Application to model predictive control. *Automatica*, 33(3):463–470, 1997.

[44] J. J. E. Slotine, J. K. Hedrick, and E. A. Misawa. On sliding observers for nonlinear systems. *Journal of Dynamic Systems, Measurement, and Control*, 109(3):245–252, 1987.

[45] Yongkyu Song. *Estimation and Control in Discrete-Time Nonlinear Systems.* PhD thesis, University of Michigan, 1992.

[46] Yongkyu Song and Jessey W. Grizzle. The extended kalman filter as a local asymptotic observer for discrete-time nonlinear systems. *Journal of Mathematical Systems, Estimation and Control*, 5(1):59–78, 1995.

[47] Robert F. Stengel. *Stochastic Optimal Control: Theory and Application.* John Wiley and Sons, New York, 1986.

[48] M. L. Tyler and M. Morari. Estimation of cross directional properties: Scanning versus stationary sensors. *AIChE Journal*, 41(4):846–854, April 1995.

[49] Matthew L. Tyler and Manfred Morari. Stability of constrained moving horizon estimation schemes. Submitted for Publication, 1997.

[50] Mathukumalli Vidyasagar. *Nonlinear Systems Analysis.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2nd edition, 1993.

[51] N. Wiener. *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* John Wiley & Sons, Inc., New York, N. Y., 1949.

[52] Xiao-Hua Xia and Wei-Bin Gao. Nonlinear observer design by observer error linearization. *SIAM Journal on Control and Optimization*, 27(1):199–216, 1989.

Vita

Peter Klaus Findeisen was born in Nürtingen, Germany on February 5, 1970, as the second son of Konrad Findeisen and Renate Findeisen.
He graduated from Max Eyth Technisches Gymnasium, Kirchheim unter Teck, Germany in 1990. After this he fulfilled his compulsory military service as an electronic repair soldier in the army base Großengstingen . He entered the University of Stuttgart in the fall of 1991 and received his Vordiplom in Technische Kybernetik in the summer of 1993. He studied for four more semesters on his Hauptdiplom of Technische Kybernetik at the University of Stuttgart before he left for an integrated exchange program with the University of Wisconsin-Madison in Fall 1995.

Permanent Address: Peter Findeisen
Siemensstraße 10
D-72622 Nürtingen
Germany

This thesis was prepared with LaTeX 2_ε[1] by the author.

[1]This particular University of Wisconsin complient style was carved from The University of Texas at Austin styles as written by Dinesh Das (LaTeX 2_ε), Khe–Sing The (LaTeX), and John Eaton (LaTeX). Knives and chisels wielded by John Campbell and Rock Matthews.