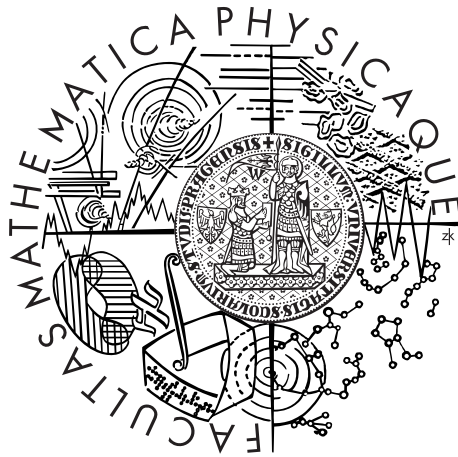


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Karel Musil

Porovnání metod pro odhad omezených veličin s aplikací na ekonomická data

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Lenka Pavelková, Ph.D.

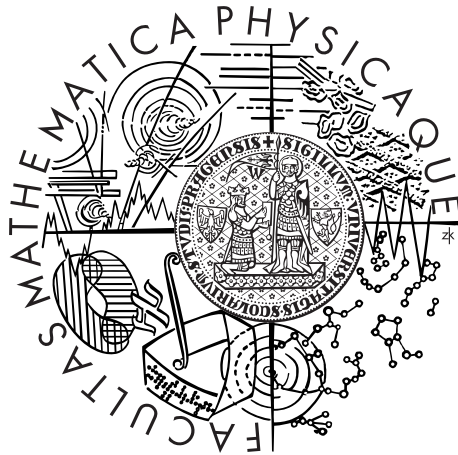
Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2013

Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Karel Musil

Comparison of Methods for Estimation of Bounded Quantities with Application to Economic Data

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: Ing. Lenka Pavelková, Ph.D.

Study programme: Mathematics

Specialization: Probability, Mathematical
Statistics and Econometrics

Prague 2013

I would like to thank Ing. Lenka Pavelková, Ph.D. who supervised, advised and supported me throughout the realization of this thesis. Furthermore, I want to owe special thanks to prof. RNDr. Tomáš Cipra, DrSc. for his assistance and valuable comments. Finally, many thanks to all those who helped me in this work.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

Prague
April 11, 2013

.....
signature of the author

Title: Comparison of Methods for Estimation of Bounded Quantities with Application to Economic Data

Author: Karel Musil

Department: Department of Probability and Mathematical Statistics

Supervisor: Ing. Lenka Pavelková, Ph.D., Institute of Information Theory and Automation of the ASCR, v.v.i.

Abstract: The thesis introduces an overview of techniques for filtering of unobserved variables using a state-space representation of a model and state inequality constraints. It is mainly aimed at a derivation of the linear Kalman filter, its extension into a form of a non-linear filter and imposing state constraints. The state uniform model with noise bounds and the sequential importance sampling, as a method of particle filters using Monte Carlo simulations, are described as alternative methods. These three methods are applied on a simple semi-structural model for a monetary policy analysis. The filtration is based on Czech macroeconomic data and reflects an imposed non-negative state constraint on the interest rate. Results of the algorithms are compared and discussed.

Keywords: State-Space Model, State Constraints, Kalman Filter, Bounded Noise, Particle Filter, Method Comparison.

Název práce: Porovnání metod pro odhad omezených veličin s aplikací na ekonomická data

Autor: Karel Musil

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Lenka Pavelková, Ph.D., Ústav teorie informace a automatizace AV ČR, v.v.i.

Abstrakt: Diplomová práce představuje přehled základních technik pro filtrování nepozorovaných proměnných při použití stavové reprezentace modelu a stavových omezení ve tvaru nerovnic. Zabývá se především odvozením Kalmanova filtru, jeho rozšířením do nelineární podoby a použitím omezení na stavové proměnné. Alternativní přístupy pomocí stavového modelu s rovnoměrně rozloženým šumem a Sequential importance sampling jako jedna z metod Particle filtrů využívající Monte Carlo simulace jsou také popsány. Všechny tři metody jsou aplikovány na semistrukturální model použitelný pro analýzu měnové politiky. Filtrace používá makroekonomická data české ekonomiky a zohledňuje nezáporné omezení úrokových sazeb jako jeden z modelových stavů. Výsledky jsou navzájem srovnány a diskutovány.

Klíčová slova: stavový model, stavové omezení, Kalmanův filtr, omezený šum, Particle filtr, porovnání metod.

Contents

List of Abbreviations	3
List of Figures	7
List of Tables	7
Introduction	9
1 Overview of Approaches and Algorithms	11
1.1 State-Space Representation	11
1.1.1 Linear State-Space Representation	12
1.1.2 Generalization of the Linear State-Space Representation	14
1.1.3 Non-Linear State-Space Representation	15
1.2 State Constraints	15
1.3 Bayesian Statistics	17
1.3.1 Bayes' Rule	17
1.3.2 Bayesian Filtering	19
2 The Linear Kalman Filter and Linear Constraints	21
2.1 The Kalman Filter Derivation	21
2.1.1 The Kalman Filter: Initialization	22
2.1.2 The Kalman Filter: Forecasting and Updating	22
2.1.3 The Kalman Filter: Smoothing	26
2.1.4 The Kalman Filter: Set of Equations	29
2.2 Maximum Likelihood Estimation	31
2.2.1 Maximum Likelihood Estimation of Parameters	31
2.2.2 Asymptotic Properties	32
2.3 The Kalman Filter Alterations	32
2.3.1 The Kalman Filter with Non-Normal Disturbances	33
2.3.2 Time-Varying Coefficients	33
2.4 Linear Constraints	34
2.4.1 Estimate Projection	36
2.4.2 Kalman Gain Projection	40
2.4.3 Probability Density Function Truncation	41
2.4.4 State Prediction Constraints	43
2.4.5 Soft Constraints	44
3 The Non-Linear Kalman Filter and/or Non-Linear Constraints	47
3.1 Non-Linear Kalman Filters	47
3.1.1 The Extended Kalman Filter	47
3.1.2 The Unscented Kalman Filter	48
3.1.3 Modifications of the Unscented Kalman Filter	49
3.2 Constraints	50
3.2.1 Second-Order Expansion	52
3.2.2 Moving Horizon Estimation	52
3.2.3 Interior Point Method	52

4	State Uniform Model Filtration	53
4.1	State-Space Model with Bounded Noise	53
4.2	Linear State and Noise Boundary Estimation	54
4.3	Non-Linear State-Space Model with Missing Data	57
4.4	Non-Linear State and Noise Boundary Estimation	58
5	Particle Filters	61
5.1	Sequential Importance Sampling	62
5.2	Constrained Sequential Importance Sampling	65
6	Application: Economic Data Filtering	67
6.1	A Monetary Policy Semi-Structural Model	67
6.1.1	State-Space Representation of the Model	69
6.1.2	Data	70
6.1.3	Economic Interpretation	72
6.2	Imposing an Inequality Constraint	76
6.2.1	Probability Density Function Truncation	78
6.2.2	State Uniform Model	79
6.2.3	Sequential Importance Sampling	79
6.3	Comparison of Algorithms	82
6.4	Discussion of Results	88
	Conclusion	91
	Bibliography	95
	Appendix	103
	Appendix 1: Model Code in Iris	103
	Appendix 2: Matlab Code for PDF Truncation	104
	Appendix 3: Matlab Code for SU Model Filtration	105
	Appendix 4: Matlab Code for SIS Particle Filtering	107

List of Abbreviations and Notations

\equiv	Identity by definition
\approx	Approximation
\otimes	Kronecker product, defined as $\mathbf{A}_{(a,b)} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & \cdots & A_{1b}\mathbf{B} \\ \vdots & & \vdots \\ A_{a1}\mathbf{B} & \vdots & A_{ab}\mathbf{B} \end{bmatrix}$
\cap	Set intersection
\in	Element of a set
\propto	Equality up to a constant factor (proportionality)
$\ \cdot\ $	Euclidean norm
∂	Partial derivative of a function with respect to one of its variable
$\mathbf{0}, \mathbf{1}$	Column vector or matrix of zeros and ones respectively (index is omitted when the size is obvious from context)
\mathbf{A}, \mathbf{B}	Matrix A, B (index is omitted when the size is obvious). The following notation is reserved: $\mathbf{A}, \mathbf{H}, \mathbf{F}$ Matrices of a state-space model representation \mathbf{I} Square identity matrix $\mathbf{K}_t, \dot{\mathbf{K}}_t$ (Kalman) gain matrix at time t $\mathbf{P}_{\cdot \cdot}$ MSE matrix for inferences about a state vector \mathbf{Q}, \mathbf{R} Variance-covariance matrices of disturbances in a model
\mathbf{A}^\top	Transposition of matrix \mathbf{A}
\mathbf{A}^{-1}	Inversion of matrix \mathbf{A}
$\mathbf{A}^{-\top}$	Transposed inversion of matrix \mathbf{A}
$ \mathbf{A} $	Determinant of square matrix \mathbf{A}
\mathbf{a}, \mathbf{b}	Column vector a, b (index is omitted when the size is obvious)
\mathbf{a}_t	Value of \mathbf{a} in discrete time instant t ($t = 1, \dots, T$). The following notation is reserved: \mathbf{x}_t State vector at time t \mathbf{y}_t Observed (measurement, output) vector at time t \mathbf{u}_t Exogenous vector at time t \mathbf{v}_t State disturbance (error) at time t \mathbf{w}_t Measurement error at time t
$\{\mathbf{a}_k\}_{k=1}^K, \mathbf{a}\mathbf{a}_K$	Ordered sequence (collection) of variables (vectors, matrices, vector functions) for $k = 1, \dots, K$, $\mathbf{a}\mathbf{a}_K = \{\mathbf{a}_k\}_{k=1}^K \equiv \{\mathbf{a}_1^T, \dots, \mathbf{a}_K^T\}$
$\mathbf{a}\mathbf{a}_K^{l:m}$	Sequence (collection) of variables (vectors, matrices, vector functions), where $1 < l < m < K$, $\mathbf{a}\mathbf{a}_K^{l:m} \equiv \{\mathbf{a}_1^T, \dots, \mathbf{a}_l^T, \mathbf{a}_m^T, \dots, \mathbf{a}_K^T\}$
$\bar{\mathbf{a}}, \underline{\mathbf{a}}$	Upper and lower bound on \mathbf{a} , meant entrywise
$\hat{\mathbf{a}}_{t+s t}$	Linear projection of \mathbf{a}_{t+s} on a set of variables at time t
$\hat{\mathbf{a}}_t$	Point estimate of \mathbf{a}_t
$\mathbf{a}(\cdot)$	Vector function of a specified dimension
$\mathbf{a}'(\cdot), \mathbf{a}''(\cdot)$	Matrix of the first and second partial derivatives of function $\mathbf{a}(\cdot)$
AR(1)	Autoregression of the first order
$\operatorname{argmax}(\cdot)$	Argument of maximum

$\operatorname{argmin}(\cdot)$	Argument of minimum
CIUKF	Constrained interval unscented Kalman filter
$\operatorname{cov}(\cdot)$	Covariance
CPI	Consumer Price Index
CUKF	Constrained unscented Kalman filter
$\mathbb{E}(\cdot)$	Expectation
$\mathbb{E}(\cdot \cdot)$	Conditional mean value
\mathbf{e}_i	The i -th canonical vector
ECUKF	Equality constrained extended Kalman filter
EKF	Extended Kalman filter
EP	Estimate projection
$\exp\{\cdot\}$	Number e (the base for the natural logarithm) raised to a power
GDP	Gross domestic product
GP	Gain projection
$\mathbb{I}(\cdot)$	Indicator function
i.i.d.	Independently and identically distributed variable
IPLM	Interior point likelihood maximization
IUKF	Interval unscented Kalman filter
KF	Kalman filter
$\mathcal{L}_a(\cdot), \mathcal{R}_a(\cdot)$	Operators adding zero columns of a length to the specified matrix from the left or right side
$\lim(\cdot)$	Limit
$\log(\cdot)$	Natural logarithm
MAP est.	Maximum a posteriori estimation
MC	Monte Carlo
MHE	Moving horizon estimation
MR	Model reduction
MSE	Mean squared error
\mathbb{N}	Positive integers
\mathbb{N}_0	Non-negative integers (positive integers including zero)
$\mathcal{N}(\cdot, \cdot)$	(Multidimensional) Normal distribution
$\max(\cdot)$	Maximum of values
$\min(\cdot)$	Minimum of values
QoQ ann.	Quarter-to-quarter annualized (change)
PDF, $p(\cdot \cdot)$	(Conditional) Probability density function
PDFT	Probability density function truncation
PF	Particle filter
PM	Perfect measurement
PME_t^i	Predicted measurement (variable) error of the i -th variable at time t
PUKF	Projected unscented Kalman filter
\mathbb{R}	Real numbers
\mathbb{R}^n	The set consisting of all real n -dimensional vectors
RMSE, RMSE^i	Root mean square error (for the i -th variable)
SCKF	Smoothly constrained Kalman filter
SIS	Sequential importance sampling
SP	System projection
t, τ	Time indices (discrete, $t = 1, \dots, T$)
T	Number of dates included in a sample ($T < \infty$)

TIUKF	Truncated interval unscented Kalman filter
$\text{Tr}(\cdot)$	Trace of matrix
TUKF	Truncated unscented Kalman filter
UKF	Unscented Kalman filter
UT	Unscented transformation
SU model	State uniform model
$\mathcal{U}(\cdot, \cdot)$	Uniform distribution (on multivariate boxes)
URNDDR	Unscented recursive non-linear dynamic data reconciliation
$\text{var}(\cdot)$	Variance
2E	Second-order expansion of non-linear constraints
2UKF	Two-step unscented Kalman filter

By the notation, it is implied that all vectors and matrices have compatible dimensions, which are omitted for simplicity.

All integrals used are definite and multivariate ones. The integration domain coincides with the support of the PDF in its argument.

List of Figures

Figure	Title	Page
1.1	Application of a Filter	11
1.2	Method Overview of the Constrained State Estimation	17
2.1	Constrained Filter Approach	35
2.2	Principle of the Projection Approach	36
2.3	Unconstrained and Constrained PDF - The Unconstrained Estimate Does Not Satisfy the Constraint	42
2.4	Unconstrained and Constrained PDF - The Unconstrained Estimate Satisfies the Constraints	42
3.1	Errors in Linear Approximation of Nonlinear Constraints	51
5.1	Sequential Importance Sampling Algorithm	64
6.1	Seasonal Adjustment of Data	71
6.2	Measured Variables of Model	72
6.3	Measurement Errors after Filtration	74
6.4	Trends after Filtration	75
6.5	Structural Variables Decomposition after Filtration	76
6.6	Neutral Nominal and Real Interest Rate Trend Filtration	77
6.7	Kalman Filter: Smoothed and Predicted Values of Measurement Variables and Real Interest Rate Trend	78
6.8	PDF Truncation: Results	80
6.9	SU Model: Results	81
6.10	SU Model: Noise Boundaries for Measurement Variables	82
6.11	SIS: Results	83
6.12	Comparison of Predicted Measurement Error	86
6.13	Comparison of Real Interest Rate Trend	87

List of Tables

Table	Title	Page
6.1	Calibration of the Model	73
6.2	Comparison of Predicted Measurement Errors	84

Introduction

A dynamic estimation problem has received an attention over past decades and many techniques have been introduced to solve it. However, these methods, especially in case of non-linear systems, usually ignore constraints or use some approximations of the model to handle this problem. This can result in inaccurate results.

The thesis introduces several algorithms to estimate a constrained dynamic system. Constraints are only in a form of inequalities. Three of them are described in details and subsequently applied to an economic data. This enables to practically compare them with respect to their speed, accuracy, reliability and initiation.

The thesis also briefly overviews most of the currently available approaches that can be practically applied to solve the dynamic systems and offer a way how to avoid a shortcoming of either ignoring or simplifying several aspects of the initial dynamic estimation - not only with respect to the constraints but also to a non-Gaussian processes. Additionally, the thesis offers a literature survey about different filtering techniques and relatively detailed description of the basics behind fundamental filtering techniques - the Kalman filter, filtering with uniformly distributed and bounded noise and particle filters.

One of a commonly used method is the Kalman filter.¹ Based on [88], the Kalman filter is a set of mathematical equations that provides an efficient computational (recursive²) means to estimate the state of a process, in a way that minimizes the mean of squared errors. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. Additionally, the Kalman filter is possible to specify as the minimum-variance state estimator for linear dynamic systems with Gaussian noise, see [68]. Moreover, the Kalman filter is the minimum-variance linear state estimator for linear dynamic systems with non-Gaussian noise, see [72]. However, for non-linear systems it is not possible to find an optimal state estimator in a closed form, but some modifications must be employed.

The idea behind the Kalman filter is to compute only the mean and the covariance statistics of variables of our interest. This has the advantage of being completely tractable. The advantageous power of the filter is that it operates on-line which implies that to compute the best estimate of the state and its uncertainty, only an update of the previous estimates by the new measurement is required. This implies that it is not necessary to consider all previous data again to compute the optimal estimates. Due to this tractability, the method passed several modification and extensions.

In case that we have additional information about the system - e.g. the system satisfies equality and/or inequality constraints - we should choose an approach not ignoring or reducing this information, but fully and correctly incorporating

¹The Kalman filter was originally introduced by Rudolf Kalman in his famous publication [40] using the construction of the state estimation filter based on probability theory, and more specifically, on the properties of conditional Gaussian random variables.

²The term recursive means that the estimation is running on-line and the estimates are permanently refined.

it. In this situation the additional information are supposed to improve results and not only marginally.

If both the system and state constraints are linear, an optimal constrained linear state estimate (measured by its minimum-variance) of various methods results in the same state estimate. If the system or/and constraints are non-linear, then constrained filtering is not optimal in general, and different approaches give different results, for a further discussion see [73]. This increases an interest in this area, because these problems are not only theoretical, but also practical - the state constrained systems include among others navigations [75], camera tracking [37], fault diagnostic [74], robotics [77], chemical processes [85] and biomedical systems [17]. Generally, [89] concludes that the Kalman filter in practice can be used for almost everything that moves or has some dynamics.

Noises of a system describing behavior of unobserved (state) and observed evolution are often supposed to have normal distribution and the problem is then solved by means of the Kalman filtering. However, the unbounded support of the Gaussian distribution can cause difficulties in case when the estimated quantity is physically restricted.

One of a presented method in this thesis is an algorithm for a state estimation of the discrete-time dynamic system which uncertainties are bounded. It uses Bayesian approach and evaluates maximum a posteriori probability estimates of states and parameters. As the model uncertainties are supposed to have a bounded support, the searched estimates lie within an area that is described by the system of inequalities.

A third applicable approach is Monte-Carlo sampling alias particle filtering which is an estimation technique based on an idea that inequality constraints are imposed by accept/reject steps in the algorithm. The appropriate constrained prior distribution is truncated or modified to satisfy the constraints. This ensures that the posterior also satisfies the constraints. The Monte-Carlo methods are based on simulation. Therefore, a huge amount of data is required to obtain acceptable results. The particle filtering technique will be introduced in this thesis in more details.

Although an area of filtering with state constraints is relatively wide, the thesis aims only at the fundamental approaches. To cover the aim of the thesis the following structure is employed. After a general overview, containing a description of a state-space representation of a dynamic system and basics of the Bayesian statistics, there is an introduction into the linear Kalman filter and linear state constraints as a starting point for further extensions. Next chapter covers various modifications to handle either the non-linear state-space model or/and non-linear constraints problem. Next two chapters present state uniform model filtration with bounded noises and particle filters. The second part of the thesis covers practical implementations of the algorithms on real economic data resulting in their mutual comparison - especially their speed, accuracy, reliability and challenges during initializing of each algorithm. The last part summarizes and concludes.

The practical part introduces a simple economic model applied to Czech economy data. Matlab and Iris toolbox are used for calculations. Pivotal codes are open functions and are enclosed in appendices. All the codes for generating results in this thesis are available upon request.

1. Overview of Approaches and Algorithms

This chapter offers an introduction and overview of fundamental principles and methods used throughout the rest of the thesis. After a detailed description of the state-space representation of a dynamic system, there is an introduction to state constraints related to the state-space representation. The last part covers basics of Bayesian filtering, which is used mainly in Chapter 4 and 5. Although the linear Kalman filter is a special case of Bayesian filtering under the linear, quadratic and Gaussian assumptions, it is derived by an alternative method in the next chapter.

1.1 State-Space Representation

Figure 1.1 illustrates a general application context in which a filter (typically the Kalman filter) is used. A system is driven by a set of external inputs and controls and its outputs are evaluated by measurements, such that the knowledge about the system behavior is solely given by the inputs and observed outputs. The observations usually convey some errors and uncertainties in this process. For a general discussion see [67], for further comments about a theory of systems and controlling see e.g. [35] and [78].

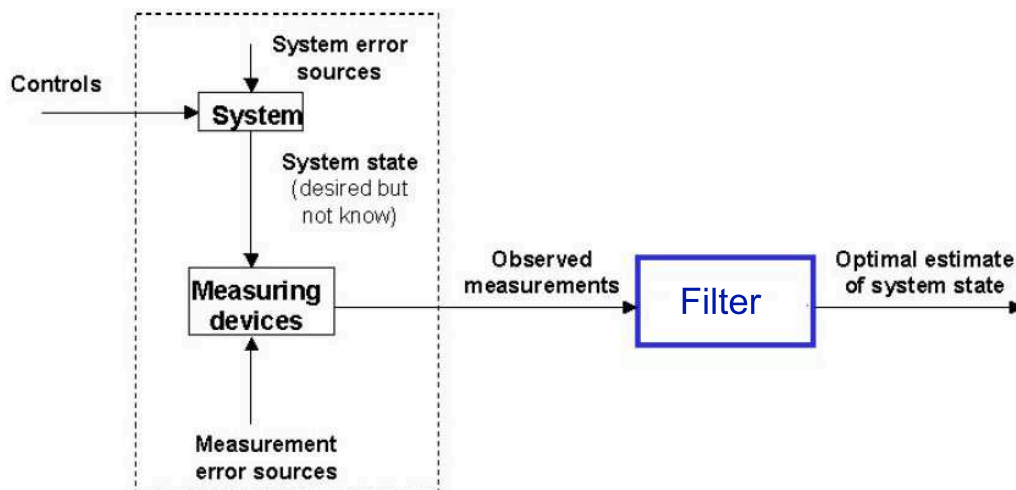


Figure 1.1: Application of a Filter (Source: [67] and author's modification)

Although dynamic systems and control theories are popular and frequently used, there are three basic reasons why they do not provide a totally sufficient means of performing this analysis, see [57]

- no mathematical system model is perfect and any model depicts only characteristics of direct interest and are an approximation,
- dynamic systems are driven not only by control inputs, but also by disturbances, which are neither controlled nor modeled deterministically, and

- no perfect and complete data about the system are provided.

A dynamic system is usually captured by a dynamic model. Many dynamic models can be rewritten in a form of a state-space form. This form enables to do an extremely simple analysis of the dynamics of a model, make forecast or evaluate the likelihood function. The state-space framework can also be used as a parsimonious time-series description of an observed vector of variables, see e.g. [34].

1.1.1 Linear State-Space Representation

A state-space representation of a more complicated linear system captures the dynamics of an observed ($n \times 1$) vector \mathbf{y}_t at time $t = 1, \dots, T \in \mathbb{N}$, in terms of a possibly unobserved ($r \times 1$) vector \mathbf{x}_t - the state vector for the system. The dynamics of the state vector are taken to be a vector generalization of a first-order autoregression. In other words, the dynamics of the state vector is governed by a Markov process, i.e. the state at \mathbf{x}_{t+1} is independent of all other states, given \mathbf{x}_t ,

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{v}_{t+1}, \quad (1.1)$$

where \mathbf{F} denotes an ($r \times r$) matrix and the ($r \times 1$) vector \mathbf{v}_t , is taken to be i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{Q})$. The previous equation is sometimes shifted backward one period, $\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{v}_t$. From the practical point of view it makes only a little difference, see [34].

A definition of the state vector \mathbf{x}_t for any particular model is determined by its construction. Their elements may or may not be identifiable with components which have an interpretation (e.g. a trend). From the technical point of view, the aim of the state-space formulation is to set up \mathbf{x}_t in such a way that it contains all the relevant information on the system at time t and that it does so by having as small a number of elements as possible. A state-space representation which reflects this fact is said to be a minimal realization. However, it does not imply that there is a unique representation for any particular problem. In fact a unique representation is an exception rather than the rule, see [34].

The observed variables are related to the state vector by the following measurement equation

$$\mathbf{y}_t = \mathbf{A}\mathbf{u}_t + \mathbf{H}\mathbf{x}_t + \mathbf{w}_t. \quad (1.2)$$

Here \mathbf{H} is an ($n \times r$) matrix of coefficients, and \mathbf{w}_t , is an ($n \times 1$) vector named as measurement error. It is assumed to be i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{R})$ and independent of \mathbf{x}_t and \mathbf{v}_t for $t = 1, 2, \dots, T$. Equation (1.2) also includes \mathbf{u}_t , a ($k \times 1$) vector of observed variables that are exogenous (predetermined) entering the equation through the ($n \times k$) matrix of coefficients \mathbf{A} , where $k, n, r \in \mathbb{N}$.¹ There is a choice if a variable is defined to be a component of the state vector, or the exogenous vector \mathbf{u}_t .²

The state equation (1.1) and observation (measurement) equation (1.2) constitute a linear state-space representation for the dynamic behavior of $\mathbf{y}_T \equiv$

¹The state-space representation of the model can be defined by using system matrices with or without transposition, check e.g. [32] or [72]. It represents an equivalent system, only the Kalman filter equations require recalculations.

²The statement that \mathbf{u}_t is predetermined or exogenous means that \mathbf{u}_t provides no information about \mathbf{x}_{t+s} or \mathbf{w}_{t+s} for $s = 1, 2, \dots$ beyond that contained in $\mathbf{y}_{t-1}, \dots, \mathbf{y}_1$.

$\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T\}$. As a result, the system is linear and any value of \mathbf{y}_t can be expressed as a linear combination of present and past of \mathbf{u}_t , \mathbf{v}_t , \mathbf{w}_{t+1} and initial state vector \mathbf{x}_1 .

Regarding the disturbances \mathbf{v}_t and \mathbf{w}_t , they are white noise and assumed to be uncorrelated at all lags

$$\mathbb{E}(\mathbf{v}_t \mathbf{v}_\tau^\top) = \begin{cases} \mathbf{Q} & \text{for } t = \tau, \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad (1.3)$$

$$\mathbb{E}(\mathbf{w}_t \mathbf{w}_\tau^\top) = \begin{cases} \mathbf{R} & \text{for } t = \tau, \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad (1.4)$$

$$\mathbb{E}(\mathbf{v}_t \mathbf{w}_\tau^\top) = \mathbf{0} \quad \text{for all } t \text{ and } \tau. \quad (1.5)$$

The state-space system is typically used to describe a finite series of observations $\{\mathbf{y}_1, \dots, \mathbf{y}_T\}$ for which assumptions about the initial value of the state vector \mathbf{x}_1 is needed. It is also assumed that \mathbf{x}_1 is uncorrelated with any realization of \mathbf{v}_t or \mathbf{w}_t for all t

$$\mathbb{E}(\mathbf{v}_t \mathbf{x}_1^\top) = \mathbf{0}, \quad \mathbb{E}(\mathbf{w}_t \mathbf{x}_1^\top) = \mathbf{0}. \quad (1.6)$$

The introduced system is quite flexible and straightforward to generalize the results to a system where \mathbf{v}_t is correlated with \mathbf{w}_t , see e.g. [4]. Moreover, the state equation (1.1) implies that \mathbf{x}_t can be rewritten as a linear function of $(\mathbf{x}_1, \mathbf{v}_2, \dots, \mathbf{v}_t)$ for $t = 2, \dots, T$

$$\mathbf{x}_t = \mathbf{v}_t + \mathbf{F}\mathbf{v}_{t-1} + \mathbf{F}^2\mathbf{v}_{t-2} + \dots + \mathbf{F}^{t-2}\mathbf{v}_2 + \mathbf{F}^{t-1}\mathbf{x}_1. \quad (1.7)$$

This together with the previous assumptions imply following

$$\mathbb{E}(\mathbf{v}_t \mathbf{x}_\tau^\top) = \mathbf{0} \quad \text{for } \tau = 1, \dots, t-1, \quad (1.8)$$

$$\mathbb{E}(\mathbf{w}_t \mathbf{x}_\tau^\top) = \mathbf{0} \quad \text{for } \tau = 1, \dots, T, \quad (1.9)$$

$$\mathbb{E}(\mathbf{w}_t \mathbf{x}_\tau^\top) = \mathbb{E}[\mathbf{w}_t(\mathbf{A}\mathbf{u}_\tau + \mathbf{H}\mathbf{x}_\tau + \mathbf{w}_\tau)] = \mathbf{0} \quad \text{for } \tau = 1, \dots, t-1, \quad (1.10)$$

$$\mathbb{E}(\mathbf{v}_t \mathbf{y}_\tau^\top) = \mathbf{0} \quad \text{for } \tau = 1, \dots, t-1. \quad (1.11)$$

Note also that when \mathbf{u}_t is deterministic, the state vector \mathbf{x}_t summarizes everything in the past that is relevant for determining future values of \mathbf{y}_{t+m} , $m \in \mathbb{N}_0$,

$$\begin{aligned} & \mathbb{E}(\mathbf{y}_{t+m} | \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{y}_t, \mathbf{y}_{t-1}, \dots) \\ &= \mathbb{E}[(\mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\mathbf{x}_{t+m} + \mathbf{w}_{t+m} | \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{y}_t, \mathbf{y}_{t-1}, \dots)] \\ &= \mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\mathbb{E}(\mathbf{x}_{t+m} + \mathbf{w}_{t+m} | \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{y}_t, \mathbf{y}_{t-1}, \dots) \\ &= \mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\mathbf{F}^m \mathbf{x}_t. \end{aligned} \quad (1.12)$$

Having the state-space representation of a model, the general problem is to find an estimate $\hat{\mathbf{x}}_{t+1}$ of \mathbf{x}_{t+1} given the measurements $\mathbf{y}_t = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$ using the state-space representation of the model.

System Stability. The system is stable, for any initial state \mathbf{x}_1 , if the state vector converges to an equilibrium solution, when \mathbf{u}_t is constant for all t . The necessary and sufficient condition for stability is that all characteristic roots of the transition matrix \mathbf{F} lays inside a unit circle. More generally, the key properties of the state-space representation are those of observability, controllability, detectability and stabilizability, see [34]. When applied to the system they provide

the basis for assessing statistical properties. In case of the control engineering, they have their meaning and interpretation.

Used Form. To summarize, the constant-parameter linear state-space model has the following specification for the rest of the thesis

- State (transition) equation

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{v}_{t+1}, \quad (1.13)$$

$$\mathbb{E}(\mathbf{v}_{t+1}\mathbf{v}_{t+1}^\top) = \mathbf{Q}. \quad (1.14)$$

- Observation (measurement, output) equation

$$\mathbf{y}_t = \mathbf{A}\mathbf{u}_t + \mathbf{H}\mathbf{x}_t + \mathbf{w}_t, \quad (1.15)$$

$$\mathbb{E}(\mathbf{w}_t\mathbf{w}_t^\top) = \mathbf{R}. \quad (1.16)$$

Writing a model in the state-space form means imposing certain values on some of the elements of $\mathbf{F}_{(r,r)}$, $\mathbf{Q}_{(r,r)}$, $\mathbf{A}_{(n,k)}$, $\mathbf{H}_{(n,r)}$ and $\mathbf{R}_{(n,n)}$, and interpreting the other elements as particular parameters of interest with respect to the $(r \times 1)$ type state vector \mathbf{x}_{t+1} , $(n \times 1)$ type observed (measurement) vector \mathbf{y}_{t+1} , $(k \times 1)$ type exogenous vector \mathbf{u}_t , $(r \times 1)$ type state discrepancy \mathbf{v}_{t+1} and $(n \times 1)$ type measurement error \mathbf{w}_t , for $t = 1, \dots, T$, where $k, n, r, T \in \mathbb{N}$.

The state-space framework can be further generalized to allow its extension, time-varying coefficient matrices, non-normal disturbances and non-linear dynamics.

1.1.2 Generalization of the Linear State-Space Representation

The previously introduced state-space representation of the model is easy to extend to include a drift and external inputs. The drift is a constant change expressed by adding $(r \times 1)$ type vector \mathbf{b} to the state equation (1.13)

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{b} + \mathbf{v}_{t+1} = \dot{\mathbf{F}}\dot{\mathbf{x}}_t + \mathbf{v}_{t+1}, \quad (1.17)$$

where the constant is incorporated through the redefinitions of $\dot{\mathbf{F}} = [\mathbf{F}, \mathbf{b}]$ and $\dot{\mathbf{x}}_t = [\mathbf{x}_t, \mathbf{1}]$. For the external inputs, it is possible to use an extension of the exogenous vector \mathbf{u}_t

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{v}_{t+1}. \quad (1.18)$$

This model is used when we want to control the system. This generalization works in the same way as the original state-space representation introduced earlier and it is not used later because of a simplification of further calculations.

In practice, the process noise covariance \mathbf{Q} and measurement noise covariance \mathbf{R} matrices might change with each time step or measurement, however, here we assume they are constant.

To conclude this section and to demonstrate a flexibility of the linear state-space representation of the dynamic system, as discussed in [63], let's generalize the state and measurement equations into the following form

$$\mathbf{x}_{t+1} = \tilde{\mathbf{B}}_t\mathbf{u}_t + \tilde{\mathbf{F}}_t\mathbf{x}_t + \tilde{\mathbf{D}}_t + \mathbf{v}_{t+1}, \quad (1.19)$$

$$\mathbf{y}_t = \tilde{\mathbf{A}}_t\mathbf{x}_t + \tilde{\mathbf{H}}_t\mathbf{u}_t + \tilde{\mathbf{C}}_t + \mathbf{w}_t, \quad (1.20)$$

where again \mathbf{x}_t , \mathbf{u}_t and \mathbf{y}_t are state, input and measurement vectors respectively. Model matrices $\tilde{\mathbf{A}}_t$, $\tilde{\mathbf{B}}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{D}}_t$, $\tilde{\mathbf{H}}_t$ and $\tilde{\mathbf{F}}_t$ are of appropriate dimensions. They are sums of the form

$$\tilde{\mathbf{A}}_t = \mathbf{A}_t + \mathbf{A}, \quad \tilde{\mathbf{B}}_t = \mathbf{B}_t + \mathbf{B}, \quad \dots, \quad (1.21)$$

where

- \mathbf{A}_t contains known, generally time-variant entries of $\tilde{\mathbf{A}}_t$, and
- \mathbf{A} contains unknown time-invariant entries of $\tilde{\mathbf{A}}_t$ and zeros.

This holds similarly for other system matrices. A discussion about a time-varying coefficients of the system matrices is covered in Subsection 2.3.2.

1.1.3 Non-Linear State-Space Representation

Let's rewrite the discrete state-space representation from Subsection 1.1.1 into a non-linear form

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t) + \mathbf{v}_{t+1} \quad (1.22)$$

$$\mathbf{y}_t = \mathbf{a}(\mathbf{u}_t) + \mathbf{h}(\mathbf{x}_t) + \mathbf{w}_t. \quad (1.23)$$

In the above equations, we see that the transition matrix \mathbf{F} has been replaced by the non-linear (real) vector-valued function $\mathbf{f} : \mathbb{R}^r \rightarrow \mathbb{R}^r$, and similarly, the matrix \mathbf{H} , which transforms a vector from the state-space into the measurement space, is replaced by the non-linear (real) vector-valued function $\mathbf{h} : \mathbb{R}^r \rightarrow \mathbb{R}^n$. As before, \mathbf{v}_t and \mathbf{w}_t are assumed to be white noise with covariance matrices \mathbf{Q} and \mathbf{R} , respectively.

The previous representation is a special case of the following one

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \mathbf{v}_{t+1}) \quad (1.24)$$

$$\mathbf{y}_t = \mathbf{h}(\mathbf{u}_t, \mathbf{x}_t, \mathbf{w}_t). \quad (1.25)$$

The state equation can be further extended by \mathbf{u}_t ,

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{u}_t, \mathbf{x}_t, \mathbf{v}_{t+1}). \quad (1.26)$$

1.2 State Constraints

The state-space formulation of a model enables to impose state constraints. The thesis discusses linear and non-linear systems extended by linear and non-linear state inequality constraints.

A linear system with linear constraints is presented first. The linear Kalman filter is introduced together with various ways of including linear constraints. These include:

- estimate projection [75],
- gain projection [31], and

- probability density function (PDF) truncation [72], [74].

Under certain conditions, all these approaches result in the same state estimate.

Further, systems which are non-linear or/and have non-linear constraints are considered. These are:

- the extended and unscented Kalman filter and their various modifications [72],
- second-order expansion of the constraints [90],
- moving horizon estimation [66], [65], and
- interior point approaches [7].

In this situation, if the system or constraints are non-linear, constraint filtering is not optimal because different methods give different results.

The state model with uniform innovations (SU model) proposes an alternative to the standardly used linear and non-linear state-space model with normal innovations that leads to the Kalman filter or its modifications (see the previously listed methods). By the SU model, the state and measurement innovations are considered to have a uniform distribution. The main advantages of the proposed method are the simplicity of the estimation algorithm and a possibility to estimate both the parameters and states including the innovation boundaries.

Another technique is based on particle filters. They operate by propagating many state estimates, called particles, and are generated by Monte Carlo simulations, which are distributed according to the PDF of a state. The sequential importance sampling, as one of a method of particle filtering, can be easily extended to incorporate imposed constraints.

Approaches to Constraints. The possible approaches to constrained state estimation reflecting a system and constraint type can be delineated by a flowchart designed by [73] and captured in Figure 1.2.

The used acronyms (in alphabetical order) and references for the methods (in case they are not stated earlier) are following CIUKF: constrained interval unscented Kalman filter [82], CUUKF: constrained unscented Kalman filter [82], ECUKF: equality constrained extended Kalman filter [81], EKF: extended Kalman filter [72], EP: estimate projection, GP: gain projection, IPLM: interior point likelihood maximization, IUKF: interval unscented Kalman filter [82], MHE: moving horizon estimation, MR: model reduction, PDFT: probability density function truncation, PF: particle filter, PM: perfect measurement, PUKF: projected unscented Kalman filter [82], SCKF: smoothly constrained Kalman filter [23], SP: system projection, TIUKF: truncated interval unscented Kalman filter [82], TUKF: truncated unscented Kalman filter [82], UKF: unscented Kalman filter, URNDDR: unscented recursive non-linear dynamic data reconciliation [84], 2E: second-order expansion of non-linear constraints, and 2UKF: two-step unscented Kalman filter [36].

The references quoted together with the method are also the main sources for description of the method further in the thesis, if it is not stated otherwise.

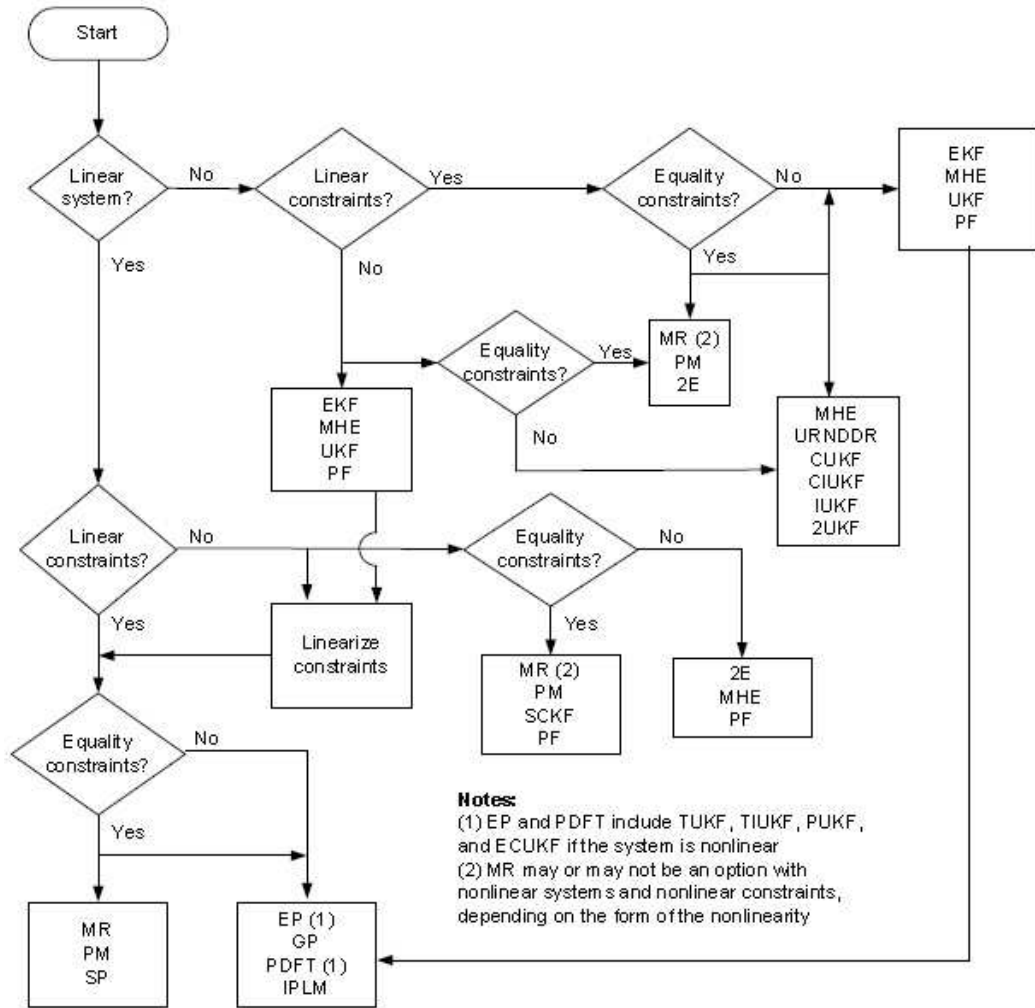


Figure 1.2: Method Overview of the Constrained State Estimation (Source: [73])

1.3 Bayesian Statistics

The state-space formulation and the requirement for the updating of information on receipt of new measurements are ideally suited for the Bayesian approach. This provides a rigorous general framework for dynamic state estimation problems.

1.3.1 Bayes' Rule

During the thesis only joint PDFs with following properties are used with the support of the PDF (a, b, c) , see e.g. [43]

- the chain rule relating conditional PDFs $p(a|b, c)$ and $p(b|c)$ and a marginal PDF $p(c)$

$$p(a, b, c) = p(a|b, c)p(b|c)p(c) \quad (1.27)$$

yielding

$$p(a, b|c) = p(a|b, c)p(b|c) = p(b|a, c)p(a|c), \quad (1.28)$$

- non-negativity

$$p(a, b|c), p(a|b, c), p(a|b) \geq 0, \quad (1.29)$$

- normalization

$$\int p(a, b|c) da db = 1 \qquad \int p(a|b, c) da = 1, \quad (1.30)$$

- marginalization of PDFs

$$p(b|c) = \int p(a, b|c) da \qquad p(a|c) = \int p(a, b|c) db. \quad (1.31)$$

The meaning of basic PDFs is as follow

- $p(a, b|c)$: joint PDF on (a, b) conditioned by c ,
- $p(a|c)$: marginal PDF on a conditioned by c (with no information on b),
- $p(b|a, c)$: marginal PDF on b conditioned by a, c .

These properties imply Bayes' rule

$$p(b|a, c) = \frac{p(a|b, c)p(b|c)}{\int p(a|b, c)p(b|c)db} = \frac{p(a|b, c)p(b|c)}{p(a|c)} \propto p(a|b, c)p(b|c). \quad (1.32)$$

Based on the Bayesian view, the system is described by probability density functions. The quantities describing the system consist generally of observable outputs $\mathbf{y}\mathbf{y}_T \equiv \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T\}$ and exogenous optional inputs $\mathbf{u}\mathbf{u}_T \equiv \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_T\}$. They together form data which is a collection of the outputs and inputs, $\mathbf{u}\mathbf{y}_T \equiv \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_T, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T\}$. Internal quantities that are never observed directly and consist of system states $\mathbf{x}\mathbf{x}_T \equiv \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$ and time invariant parameters θ . The PDF

$$p(\mathbf{u}\mathbf{y}_T, \mathbf{x}\mathbf{x}_T, \theta), \quad (1.33)$$

describing both observed and internal quantities, can be decomposed into a product of the following elements (for all t)

- observation model (the measurements \mathbf{y}_t are related to past data $\mathbf{u}\mathbf{y}_{t-1}$ and inputs \mathbf{u}_t):

$$\{p(\mathbf{y}_t|\mathbf{u}_t, \mathbf{u}\mathbf{y}_{t-1}, \mathbf{x}_t, \theta)\}_{t=1}^T, \quad (1.34)$$

- time evolution model (the evolution of \mathbf{x}_t):

$$\{p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{u}\mathbf{y}_{t-1}, \mathbf{x}_{t-1}, \theta)\}_{t=1}^T, \quad (1.35)$$

- controller (as the controller does not depend on the internal unknown quantities \mathbf{x}_t and θ , it plays no role in estimation. Therefore, the knowledge of the controller is not required and the generated input values must be known only):

$$\{p(\mathbf{u}_t|\mathbf{u}\mathbf{y}_{t-1}, \mathbf{x}_{t-1}, \theta)\}_{t=1}^T \equiv \{p(\mathbf{u}_t|\mathbf{u}\mathbf{y}_{t-1})\}_{t=1}^T, \quad (1.36)$$

- prior PDF (the initial data $\mathbf{u}\mathbf{y}_0$ coincide with the prior information about \mathbf{x}_0):

$$p(\mathbf{x}_0, \theta). \quad (1.37)$$

It holds

$$\begin{aligned}
& p(\mathbf{u}\mathbf{y}_T, \mathbf{x}\mathbf{x}_T, \theta) \\
&= p(\mathbf{x}_0, \theta) \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{u}_t, \mathbf{u}\mathbf{y}_{t-1}, \mathbf{x}_t, \theta) p(\mathbf{x}_t | \mathbf{u}_t, \mathbf{u}\mathbf{y}_{t-1}, \mathbf{x}_{t-1}, \theta) p(\mathbf{u}_t | \mathbf{u}\mathbf{y}_{t-1}) \\
&\propto p(\mathbf{x}_0, \theta) \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{u}_t, \mathbf{u}\mathbf{y}_{t-1}, \mathbf{x}_t, \theta) p(\mathbf{x}_t | \mathbf{u}_t, \mathbf{u}\mathbf{y}_{t-1}, \mathbf{x}_{t-1}, \theta). \tag{1.38}
\end{aligned}$$

The Bayesian state estimation works with characteristics of the joint PDF (1.38).

1.3.2 Bayesian Filtering

In the Bayesian approach to dynamic state estimation, one attempts to construct the posterior PDF of the state based on all available information, including the set of received measurements. Since this PDF embodies all available statistical information, it may be said to be the complete solution to the estimation problem. In principle, an optimal estimate (with respect to any criterion) of the state may be obtained from the PDF.

Using a Bayesian approach, the prediction stage uses the system model to predict the state PDF forward from one measurement time to the next. Since the state is usually subject to unknown disturbances, prediction generally translates, deforms, and spreads the state PDF. The update operation uses the latest measurement to modify the prediction PDF. This is achieved using Bayes' theorem, which is the mechanism for updating knowledge about the target state in the light of extra information from new data.

Consider the linear state space equation in the form of (1.13). From a Bayesian perspective, we would like to recursively calculate some degree of belief in the state \mathbf{x}_t at time t , taking different values, given measurements $\mathbf{y}\mathbf{y}_t$ up to time t . Thus, it is required to construct the PDF, $p(\mathbf{x}_t | \mathbf{y}\mathbf{y}_t)$, assuming the initial PDF $p(\mathbf{x}_0 | \mathbf{y}\mathbf{y}_0) \equiv p(\mathbf{x}_0)$ of the state vector, the prior, is available. Then, in principle, the PDF may be obtained, recursively, in two stages: prediction and update.

Suppose that the required PDF $p(\mathbf{x}_{t-1} | \mathbf{y}\mathbf{y}_{t-1})$ at time $t - 1$ is available. The prediction step uses the state equation of the state-space model. The prior PDF of the state at time t is obtained via the Chapman-Kolmogorov equation

$$p(\mathbf{x}_t | \mathbf{y}\mathbf{y}_{t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{y}\mathbf{y}_{t-1}) d\mathbf{x}_{t-1}. \tag{1.39}$$

In the previous equation, it was used that $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}\mathbf{y}_{t-1}) = p(\mathbf{x}_t | \mathbf{x}_{t-1})$, as (1.13) describes a Markov process of the first order. The probabilistic model of the state evolution $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ is defined by the state equation (1.13) and the known statistics of \mathbf{v}_t .

At every time step, a measurement becomes available and this may be used to update the prior via Bayes' rule

$$p(\mathbf{x}_t | \mathbf{y}\mathbf{y}_t) = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}\mathbf{y}_{t-1})}{p(\mathbf{y}_t | \mathbf{y}\mathbf{y}_{t-1})} = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}\mathbf{y}_{t-1})}{\int p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{y}_t | \mathbf{y}\mathbf{y}_{t-1}) p(\mathbf{x}_t | \mathbf{y}\mathbf{y}_{t-1})}, \tag{1.40}$$

where the normalizing constant depends on the likelihood function $p(\mathbf{y}_t | \mathbf{x}_t)$ defined by the measurement model. At the update stage (1.40), the measurement

\mathbf{y}_t is used to modify the prior density to obtain the required posterior density of the current state.

For dynamic systems, a recursive formulation of Bayesian estimation may be represented by equation (1.40) as

$$p(\mathbf{x}_t|\mathbf{y}\mathbf{y}_{t-1}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}\mathbf{y}_{t-1}), \quad (1.41)$$

where the prior, $p(\mathbf{x}_t|\mathbf{y}\mathbf{y}_{t-1})$, is combined with the most current information of the system, $p(\mathbf{y}_t|\mathbf{x}_t)$, to find the posterior, $p(\mathbf{x}_t|\mathbf{y}\mathbf{y}_{t-1})$. Each term in the previous equation may be obtained as follows. For the prior, $p(\mathbf{x}_t|\mathbf{y}\mathbf{y}_{t-1})$, is used equation (1.39). Similarly, $p(\mathbf{y}_t|\mathbf{x}_t)$ may be found as

$$p(\mathbf{y}_t|\mathbf{x}_t) = \int p(\mathbf{y}_t|\mathbf{x}_t, \nu_t)p(\nu_t|\mathbf{x}_t)d\nu_t, \quad (1.42)$$

where ν_t is known.

2. The Linear Kalman Filter and Linear Constraints

This chapter introduces an algorithm for a sequential updating linear projection of a dynamic system. The idea is to use the dynamic system in a form called the state-space representation, which was introduced in the previous chapter. It is used a linear state-space representation from Subsection 1.1.1 throughout the whole chapter if it is not stated otherwise. After the filter is derived, properties of the filter are analyzed. This chapter is based mainly on [32] and [33].

A discrete time Kalman filter construction can be based on probability theory or deterministic approach. The probabilistic approach assumes a linearity of the predictor (this formulation based on linearity is mentioned in the original Kalman's paper, see [40] and [58] or [89]). If we adopt a Bayesian viewpoint, then we want the filter to propagate the conditional probability density of the desired quantities, conditioned on knowledge of actual data coming from measurements. An alternative construction is completely deterministic and can be followed by two cases. In the first case, a criterion based on the minimization of the prediction error is used. In the second case, the criterion can be modified by adding a penalty term. For further discussion see [70]. The thesis follows the minimization of the prediction error based on mean square error criterion.

2.1 The Kalman Filter Derivation

Regarding the state-space representation (1.13) – (1.16), the Kalman filter (KF) is an algorithm for calculating an optimal forecast of the value of \mathbf{x}_t on the basis of information observed through time $t - 1$, assuming that the values of matrices \mathbf{F} , \mathbf{Q} , \mathbf{A} , \mathbf{H} and \mathbf{R} are known. If the system matrices do not change over time, the model is said to be time-invariant or time-homogeneous, see [34]. Here the discrete time Kalman filter is introduced and derived.

A motivation behind the Kalman filter algorithm is a calculation of the state vector linear forecast on the basis of data observed till time t ,

$$\hat{\mathbf{x}}_{t+1|t} = \mathbb{E}(\mathbf{x}_{t+1} | \mathbf{xy}_t), \quad (2.1)$$

where $\mathbf{xy}_t \equiv \{\mathbf{y}_t, \mathbf{y}_{t-1}, \dots, \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_1\}$ for all t . The Kalman filter calculates these forecast recursively, generating $\hat{\mathbf{x}}_{1|0}, \hat{\mathbf{x}}_{2|1}, \dots, \hat{\mathbf{x}}_{T|T-1}$. At every steps a mean squared error (MSE) is calculated, represented by the following ($r \times r$) matrix

$$\mathbf{P}_{t+1|t} \equiv \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top]. \quad (2.2)$$

Notice that the Kalman filter maintains the first two moments of the state distribution through $\hat{\mathbf{x}}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$. In other words, state estimate $\hat{\mathbf{x}}_{t+1|t}$ reflects the mean – the first moment – of the state distribution (it is normally distributed if the random variables \mathbf{v}_t and \mathbf{w}_t are assumed to be independent of each other, white and with normal probability distributions $\mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\mathcal{N}(\mathbf{0}, \mathbf{R})$ respectively). Estimate error covariance $\mathbf{P}_{t+1|t}$ reflects the variance of the state

distribution – the second central moment. This means that the probability of the $\hat{\mathbf{x}}_{t+1|t}$ conditioned on all measurements (based on Bayes’ rule) is

$$p(\mathbf{x}_{t+1}|\mathbf{xy}_t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t+1|t}, \mathbf{P}_{t+1|t}). \quad (2.3)$$

For details see [88], for more details about the probabilistic origin of the Kalman filter see e.g. [12] or [24].

2.1.1 The Kalman Filter: Initialization

The recursion starts with $\hat{\mathbf{x}}_{1|0}$, which denotes a forecast of \mathbf{x}_1 based on no observations. This is just an unconditional mean of \mathbf{x}_1 ,

$$\hat{\mathbf{x}}_{1|0} = \mathbb{E}(\mathbf{x}_1) \quad (2.4)$$

with the associated MSE

$$\mathbf{P}_{1|0} = \mathbb{E}[(\mathbf{x}_1 - \mathbb{E}\mathbf{x}_1)(\mathbf{x}_1 - \mathbb{E}\mathbf{x}_1)^\top] = \text{var } \mathbf{x}_1. \quad (2.5)$$

If the initial state \mathbf{x}_1 is not available, it is replaced with an arbitrary option of the analyst’s best guess to the initial value. In this content, $\mathbf{P}_{1|0}$ is a positive definite matrix summarizing the confidence in this guess. Higher values for the diagonal elements of $\mathbf{P}_{1|0}$ register greater uncertainty about the truth values of \mathbf{x}_1 . For the univariate models, it is possible to construct a proper prior information about \mathbf{x}_1 and $\mathbf{P}_{1|0}$ - a diffuse prior - based on the first several observations. For an example see [34]. A generalization of this algorithm offers [25] and some possible modifications can be found in e.g. [28].

To continue, the Kalman filter is an algorithm for calculating the sequence $\{\hat{\mathbf{x}}_{t+1|t}\}_{t=1}^T$ and $\{\mathbf{P}_{t+1|t}\}_{t=1}^T$, where $\hat{\mathbf{x}}_{t+1|t}$ denotes the optimal forecast of \mathbf{x}_{t+1} based on observation of \mathbf{xy}_t and $\mathbf{P}_{t+1|t}$ expresses the MSE of this forecast. An iteration of the filter starts using equations (2.4) and (2.5).

2.1.2 The Kalman Filter: Forecasting and Updating

Given the initial values $\hat{\mathbf{x}}_{1|0}$ and $\mathbf{P}_{1|0}$, the next step is to calculate the values for the next date, $\hat{\mathbf{x}}_{2|1}$ and $\mathbf{P}_{2|1}$. The calculations for $t = 2, 3, \dots, T$ have the same basic form, thus they will be described in general term for step t : $\hat{\mathbf{x}}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$ are calculated given values of $\hat{\mathbf{x}}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$.¹

Forecasting \mathbf{y}_t . Since \mathbf{w}_t is independent of \mathbf{u}_t and \mathbf{xy}_{t-1} (or equivalently $\mathbb{E}(\mathbf{x}_t|\mathbf{u}_t, \mathbf{xy}_{t-1}) = \mathbb{E}(\mathbf{x}_t|\mathbf{xy}_{t-1}) = \hat{\mathbf{x}}_{t|t-1}$), the forecast of \mathbf{y}_t conditional on \mathbf{xy}_{t-1} and \mathbf{u}_t can be inferred from the measurement equation (1.15)

$$\mathbb{E}(\mathbf{y}_t|\mathbf{u}_t, \mathbf{x}_t) = \mathbf{A}\mathbf{u}_t + \mathbf{H}\mathbf{x}_t. \quad (2.6)$$

The law of iterated projection yields

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{A}\mathbf{u}_t + \mathbf{H}\mathbb{E}(\mathbf{x}_t|\mathbf{u}_t, \mathbf{xy}_{t-1}) = \mathbf{A}\mathbf{u}_t + \mathbf{H}\hat{\mathbf{x}}_{t|t-1} \quad (2.7)$$

¹Sometimes instead of forecasting and updating steps, it is used a priori and a posteriori estimate, see [73], or predicting and correcting step, see [88]. Using the latter terminology, the algorithm can be labeled as the predictor-corrector algorithm.

and the forecast error can be rewritten

$$\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} = (\mathbf{A}\mathbf{u}_t + \mathbf{H}\mathbf{x}_t + \mathbf{w}_t) - (\mathbf{A}\mathbf{u}_t + \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) = \mathbf{H}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{w}_t \quad (2.8)$$

and has the following MSE

$$\begin{aligned} & \mathbb{E}[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})^\top | \mathbf{u}_t, \mathbf{xy}_{t-1}] \\ &= \mathbb{E}[(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})^\top] \\ &= \mathbb{E}[\mathbf{H}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\top \mathbf{H}^\top] + \mathbb{E}[\mathbf{w}_t \mathbf{w}_t^\top] \\ &= \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R}. \end{aligned} \quad (2.9)$$

The conditional variance has the previous form because $\hat{\mathbf{x}}_{t|t-1}$ is a linear function of \mathbf{xy}_{t-1} and \mathbf{w}_t is independent of both \mathbf{x}_t and $\hat{\mathbf{x}}_{t|t-1}$ (or equivalently the cross-product term disappeared because $\mathbb{E}[\mathbf{w}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\top] = \mathbf{0}$). Moreover, (1.16) and (2.2) were employed.

In the next step, an inference about the current value of \mathbf{x}_t is updated on the basis of observation of \mathbf{y}_t ,

$$\hat{\mathbf{x}}_{t|t} = \mathbb{E}(\mathbf{x}_t | \mathbf{y}_t, \mathbf{u}_t, \mathbf{xy}_{t-1}) = \mathbb{E}(\mathbf{x}_t | \mathbf{xy}_t). \quad (2.10)$$

*Updating \mathbf{x}_t .*² The conditional covariance between (2.8) and an error in forecasting the state vector is, using again $\mathbb{E}[\mathbf{w}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\top] = \mathbf{0}$ and (2.2)

$$\begin{aligned} & \mathbb{E}\{[\mathbf{x}_t - \mathbb{E}(\mathbf{x}_t | \mathbf{u}_t, \mathbf{xy}_{t-1})][\mathbf{y}_t - \mathbb{E}(\mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1})]^\top | \mathbf{u}_t, \mathbf{xy}_{t-1}\} \\ &= \mathbb{E}\{[\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}][\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}]^\top | \mathbf{u}_t, \mathbf{xy}_{t-1}\} \\ &= \mathbb{E}\{[\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}][\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}]^\top\} \\ &= \mathbb{E}\{[\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}][\mathbf{H}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{w}_t]^\top\} \\ &= \mathbb{E}\{[\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}][\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}]^\top \mathbf{H}^\top\} \\ &= \mathbf{P}_{t|t-1}\mathbf{H}^\top. \end{aligned} \quad (2.11)$$

Equation (2.11) is possible to calculate equivalently resulting in the same outcome, $\mathbb{E}\{[\mathbf{y}_t - \mathbb{E}(\mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1})][\mathbf{x}_t - \mathbb{E}(\mathbf{x}_t | \mathbf{u}_t, \mathbf{xy}_{t-1})]^\top | \mathbf{u}_t, \mathbf{xy}_{t-1}\} = \mathbf{H}\mathbf{P}_{t|t-1}$.

The distribution of the vector $(\mathbf{y}_t^\top, \mathbf{x}_t^\top)$ conditional on \mathbf{u}_t and \mathbf{xy}_{t-1} , or more precisely for $\begin{pmatrix} \mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1} \\ \mathbf{x}_t | \mathbf{u}_t, \mathbf{xy}_{t-1} \end{pmatrix}$, is

$$\mathcal{N}\left(\begin{bmatrix} \mathbf{A}\mathbf{u}_t + \mathbf{H}\hat{\mathbf{x}}_{t|t-1} \\ \hat{\mathbf{x}}_{t|t-1} \end{bmatrix}, \begin{bmatrix} \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R} & \mathbf{H}\mathbf{P}_{t|t-1} \\ \mathbf{P}_{t|t-1}\mathbf{H}^\top & \mathbf{P}_{t|t-1} \end{bmatrix}\right). \quad (2.12)$$

Based on this, it is clear that $\mathbf{x}_t | \mathbf{u}_t, \mathbf{xy}_{t-1}$ is distributed $\mathcal{N}(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$ ³ where

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \quad (2.13)$$

²Sometimes this step itself is labeled as the Kalman filtering or the Kalman-Bucy filtering in case of the continuous version of the filter, see [18].

³Let's \mathbf{z}_1 and \mathbf{z}_2 denote $(n_1 \times 1)$ and $(n_2 \times 1)$, vectors, $n_1, n_2 \in \mathbb{N}$, and they have a joint normal distribution:

$$\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}\right).$$

Then the distribution of \mathbf{z}_2 conditional on \mathbf{z}_1 is $\mathcal{N}(\mu, \Sigma)$, where $\mu = \mu_2 + \Sigma_{2,1}\Sigma_{1,1}^{-1}(\mathbf{z}_1 - \mu_1)$ and $\Sigma = \Sigma_{2,2}\Sigma_{1,1}^{-1}\Sigma_{1,2}$. Thus the optimal forecast of \mathbf{z}_2 conditional on having observed \mathbf{z}_1 is

and the MSE of the updated projection, $\mathbf{P}_{t|t} = \mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^\top]$, is

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_{t|t-1}. \quad (2.14)$$

Forecasting \mathbf{x}_{t+1} . The next step is to calculate a forecast of \mathbf{x}_{t+1} conditional on \mathbf{xy}_t . It is clear from (1.13) that $\mathbf{x}_{t+1}|\mathbf{xy}_t \sim \mathcal{N}(\hat{\mathbf{x}}_{t+1|t}, \mathbf{P}_{t+1|t})$. Or in details, the state equation (1.13) is used to calculate a forecast of \mathbf{x}_{t+1} ,

$$\hat{\mathbf{x}}_{t+1|t} = \mathbb{E}(\mathbf{x}_{t+1}|\mathbf{xy}_t) = \mathbf{F}(\mathbf{x}_t|\mathbf{xy}_t) + \mathbb{E}(\mathbf{v}_{t+1}|\mathbf{xy}_t) = \mathbf{F}\hat{\mathbf{x}}_{t|t} + \mathbf{0} = \mathbf{F}\hat{\mathbf{x}}_{t|t}. \quad (2.15)$$

Using (2.13), it is possible to continue

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t} \quad (2.16)$$

$$\begin{aligned} &= \mathbf{F}[\hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1})] \\ &= \mathbf{F}\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}). \end{aligned} \quad (2.17)$$

The coefficient matrix is known as the (Kalman) gain matrix, denoted \mathbf{K}_t . The matrix is a $(r \times n)$ type and sometimes it is also labeled as the blending factor, see [88]

$$\mathbf{K}_t \equiv \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}. \quad (2.18)$$

The difference $(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1})$ in (2.17) is called the measurement error, or the measurement residual. The residual reflects the discrepancy between the predicted measurement $\mathbf{A}\mathbf{u}_t + \mathbf{H}\hat{\mathbf{x}}_{t|t-1}$ and the actual measurement \mathbf{y}_t . A residual of zero means that the two are in complete agreement.

The MSE of the forecast can be calculated by substituting (2.13) into a combination of (2.15) and state equation (1.13) and using (1.14)

$$\begin{aligned} \mathbf{P}_{t+1|t} &= \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top] \\ &= \mathbb{E}[(\mathbf{F}\mathbf{x}_t + \mathbf{v}_{t+1} - \mathbf{F}\hat{\mathbf{x}}_{t|t})(\mathbf{F}\mathbf{x}_t + \mathbf{v}_{t+1} - \mathbf{F}\hat{\mathbf{x}}_{t|t})^\top] \\ &= \mathbf{F}\mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^\top]\mathbf{F}^\top + \mathbb{E}[\mathbf{v}_{t+1}\mathbf{v}_{t+1}^\top] \\ &= \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^\top + \mathbf{Q} \end{aligned} \quad (2.19)$$

$$= \mathbf{F}[\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_{t|t-1}]\mathbf{F}^\top + \mathbf{Q}. \quad (2.20)$$

The last equation is known as a Riccati equation.

given by

$$\mathbb{E}(\mathbf{z}_2|\mathbf{z}_1) = \mu_2 + \Sigma_{2,1}\Sigma_{1,1}^{-1}(\mathbf{z}_1 - \mu_1),$$

with Σ characterizing the MSE of this forecast

$$\mathbb{E}[(\mathbf{z}_2 - \mu)(\mathbf{z}_2 - \mu)^\top|\mathbf{z}_1] = \Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}.$$

For further discussion and a proof see properties of the multivariate normal distribution in [3].

To apply this result, suppose that the initial value of the state vector \mathbf{x}_1 of a state-space model is drawn from a normal distribution and the disturbances \mathbf{v}_t and \mathbf{w}_t are normal. Then the distribution of \mathbf{x}_t conditional on \mathbf{xy}_{t-1} turns out to be normal for all t . A mean value of this conditional distribution is represented by the $(r \times 1)$ vector $\hat{\mathbf{x}}_{t|t-1}$ and a variance of this conditional distribution is represented by the $(r \times r)$ matrix $\mathbf{P}_{t|t-1}$. Then the mean of $\hat{\mathbf{x}}_{t|t-1}$ and variance $\mathbf{P}_{t|t-1}$ is a characterization of a distribution of \mathbf{x}_t conditional on \mathbf{xy}_{t-1} for the Kalman filter iteration for step t . The output for step t (and also the input for step $t + 1$) is the mean $\hat{\mathbf{x}}_{t+1|t}$ and the variance $\mathbf{P}_{t+1|t}$ of \mathbf{x}_{t+1} conditional on \mathbf{xy}_t .

Note that the sequence $\{\mathbf{P}_{t+1|t}\}_{t=1}^T$ is not a function of data and can be evaluated without calculating the forecasts $\{\hat{\mathbf{x}}_{t+1|t}\}_{t=1}^T$. Because $\mathbf{P}_{t+1|t}$ is not a function of data, the conditional expectation of the squared forecast error is the same as its unconditional expectation,

$$\begin{aligned}\mathbf{P}_{t+1|t} &= \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top | \mathbf{xy}_t] \\ &= \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top].\end{aligned}\quad (2.21)$$

This is a consequence of having assumed normal distributions with constant variances for \mathbf{v}_t and \mathbf{w}_t .

An alternative way of recursion for $\mathbf{P}_{t|t-1}$ is sometimes useful. Subtracting the Kalman updating equation (2.17) from the state equation (1.13) and substituting the observed equation (1.15) gives

$$\begin{aligned}\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t} &= \mathbf{F}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) - \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) + \mathbf{v}_{t+1} \\ &= (\mathbf{F} - \mathbf{K}_t\mathbf{H})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) - \mathbf{K}_t\mathbf{w}_t + \mathbf{v}_{t+1}.\end{aligned}\quad (2.22)$$

Multiplying the result by its transpose, taking expectations and recalling the definition of $\mathbf{P}_{t+1|t}$ from equation (2.2) give

$$\begin{aligned}\mathbf{P}_{t+1|t} &\equiv \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top] \\ &= (\mathbf{F} - \mathbf{K}_t\mathbf{H})\mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\top](\mathbf{F}^\top - \mathbf{H}^\top\mathbf{K}_t^\top) + \mathbf{K}_t\mathbf{R}\mathbf{K}_t^\top + \mathbf{Q} \\ &= (\mathbf{F} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1}(\mathbf{F}^\top - \mathbf{H}^\top\mathbf{K}_t^\top) + \mathbf{K}_t\mathbf{R}\mathbf{K}_t^\top + \mathbf{Q}.\end{aligned}\quad (2.23)$$

The previous equation together with the definition of the Kalman gain matrix of \mathbf{K}_t in (2.18) is the equivalent expression for the Kalman filter and produces the same sequences as generated by equation (2.20).

This step closes one loop of the KF recursion.

Kalman Filter Forecasting Method. A one-period-ahead forecast of the state vector is directly given by the state equation (1.13). A m -period-ahead forecast of the vector can be calculated from the same equation by a recursive substitution

$$\mathbf{x}_{t+m} = \mathbf{F}^m\mathbf{x}_t + \mathbf{F}^{m-1}\mathbf{x}_{t+1} + \dots + \mathbf{F}^1\mathbf{x}_{t+m-1} + \mathbf{v}_{t+m}, \quad \text{for } m = 1, 2, \dots \quad (2.24)$$

The projection of \mathbf{x}_{t+m} on \mathbf{x}_t and \mathbf{xy}_t is given by

$$\mathbb{E}(\mathbf{x}_{t+m} | \mathbf{x}_t, \mathbf{xy}_t) = \mathbf{F}^m\mathbf{x}_t \quad (2.25)$$

and the law of iterated projections yields

$$\hat{\mathbf{x}}_{t+m|t} = \mathbb{E}(\mathbf{x}_{t+m} | \mathbf{xy}_t) = \mathbf{F}^m\hat{\mathbf{x}}_{t|t}. \quad (2.26)$$

An error of the m -period-ahead forecast of the state vector is calculated in the following way

$$\mathbf{x}_{t+m} - \hat{\mathbf{x}}_{t+m|t} = \mathbf{F}^m(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}) + \mathbf{F}^{m-1}\mathbf{v}_{t+1} + \dots + \mathbf{F}^1\mathbf{v}_{t+m-1} + \mathbf{v}_{t+m} \quad (2.27)$$

from which it is clear that the MSE of the forecast is

$$\mathbf{P}_{t+m|t} = \mathbf{F}^m\mathbf{P}_{t|t}(\mathbf{F}^\top)^m + \mathbf{F}^{m-1}\mathbf{Q}(\mathbf{F}^\top)^{m-1} + \dots + \mathbf{F}\mathbf{Q}\mathbf{F}^\top + \mathbf{Q}. \quad (2.28)$$

Now, the forecast of \mathbf{y}_t on the basis of \mathbf{u}_t and \mathbf{xy}_{t-1} is calculated. If \mathbf{u}_t is deterministic, it is possible to express an exact finite sample m -period-ahead forecast by using measurement equation (1.15)

$$\mathbf{y}_{t+m} = \mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\mathbf{x}_{t+m} + \mathbf{w}_{t+m}. \quad (2.29)$$

If the state vector is defined that \mathbf{u}_t is deterministic, the dynamics of any exogenous variables can be represented through \mathbf{x}_t . Applying a law of iterated expectations results in

$$\begin{aligned} \hat{\mathbf{y}}_{t+m} &\equiv \mathbb{E}(\mathbf{y}_{t+m}|\mathbf{xy}_t) = \mathbb{E}(\mathbf{y}_{t+m}|\mathbf{y}_t, \dots, \mathbf{y}_1) = \mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\mathbf{F}^m\hat{\mathbf{x}}_{t|t} \\ &= \mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\hat{\mathbf{x}}_{t+m|t}. \end{aligned} \quad (2.30)$$

The error of this forecast is

$$\begin{aligned} \mathbf{y}_{t+m} - \hat{\mathbf{y}}_{t+m|t} &= (\mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\mathbf{x}_{t+m} + \mathbf{w}_{t+m}) - (\mathbf{A}\mathbf{u}_{t+m} + \mathbf{H}\hat{\mathbf{x}}_{t+m|t}) \\ &= \mathbf{H}(\mathbf{x}_{t+m} - \hat{\mathbf{x}}_{t+m|t}) + \mathbf{w}_{t+m} \end{aligned} \quad (2.31)$$

with the MSE

$$\mathbb{E}[(\mathbf{y}_{t+m} - \hat{\mathbf{y}}_{t+m|t})(\mathbf{y}_{t+m} - \hat{\mathbf{y}}_{t+m|t})^\top] = \mathbf{H}\mathbf{P}_{t+m|t}\mathbf{H}^\top + \mathbf{R}. \quad (2.32)$$

2.1.3 The Kalman Filter: Smoothing

Now the Kalman smoother equations are derived. They allow using forward measurements to help predict the state at the current time. Smoothing estimates are run because these estimates are usually less noisy than the measurements up till the current time only.

Up to this point, the Kalman filter was motivated as an algorithm for calculating a forecast of the state vector at time t , \mathbf{x}_t , based on information available at time $t - 1$, so as a linear function of previous observations: $\hat{\mathbf{x}}_{t|t-1} = \mathbb{E}(\mathbf{x}_t|\mathbf{xy}_{t-1})$. In some applications, especially the economic ones (see e.g. the forecast and policy analysis system implemented at the Czech National Bank, check [5] and [20]), the value of the state vector is of interest in its own and the vector has a structural interpretation. A goal then might be to form an inference about the value of \mathbf{x}_t based on the full set of data collected (including observations on $\mathbf{y}_t, \mathbf{y}_{t+1}, \dots, \mathbf{y}_T, \mathbf{x}_t, \mathbf{x}_{t+1}, \dots, \mathbf{x}_T$) or through the end of the sample (date T) to help improve the inference about the historical value that the state vector took on at any particular time t in the middle of the sample. Such an inference is known as a smoothed estimate of \mathbf{x}_t , denoted $\hat{\mathbf{x}}_{t|T} = \mathbb{E}(\mathbf{x}_t|\mathbf{xy}_T)$. The MSE of this smoothed estimate is denoted $\mathbf{P}_{t|T} = \mathbb{E}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|T})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|T})^\top$. In general, $\mathbf{P}_{t|\tau}$ denotes the MSE of an estimate of \mathbf{x}_t which is based on observation of \mathbf{yy}_τ and \mathbf{xx}_τ through date τ .

Since the smoother is based on more information than the filtered estimator, it has a MSE which is smaller than that of the filtered estimator in general. The smoothed estimator exists if its elements can be estimated with the finite MSE, that is if $\mathbf{P}_{t|T}$ is bounded.

There are three smoothing algorithms in a linear model, see [34]

- fixed-point smoothing is concerned with computing smoothed estimates of the state vector at some fixed point in a time,

- fixed-lag smoothing computes smoothed estimates for a fixed delay, and
- fixed-interval smoothing is concerned with computing the full set of all smoothed estimates for a fixed span of data.

All these algorithms are recursive and closely related to the Kalman filter. The fixed-point and fixed-lag algorithms run in parallel with the Kalman filter (by augmenting the state-space model by this point estimate and applying the Kalman filter), while the fixed-interval one is a backward recursion which starts at time T .

Fixed Interval Smoothing. The fixed interval smoothing algorithm consists of a set of recursions which start with the final quantities ($\hat{\mathbf{x}}_T$ and \mathbf{P}_T) given by the Kalman filter and work backwards.

Consider the estimate of \mathbf{x}_t based on observation through date t , $\hat{\mathbf{x}}_t$. Suppose that subsequently the true value of \mathbf{x}_{t+1} is known. Using properties of the multivariate normal distribution, as it was already used in case of the updating step of the Kalman filter, the distribution of the vector $(\mathbf{x}_{t+1}^\top, \mathbf{x}_t^\top)$ conditional on \mathbf{u}_t and \mathbf{y}_{t-1} is

$$\begin{pmatrix} \mathbf{x}_{t+1} | \mathbf{u}_t, \mathbf{y}_{t-1} \\ \mathbf{x}_t | \mathbf{u}_t, \mathbf{y}_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{t+1} | \mathbf{y}_t \\ \mathbf{x}_t | \mathbf{y}_t \end{pmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \hat{\mathbf{x}}_{t+1|t} \\ \hat{\mathbf{x}}_{t|t} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{t+1|t} & \mathbf{F}\mathbf{P}_{t|t} \\ \mathbf{P}_{t|t}\mathbf{F}^\top & \mathbf{P}_{t|t} \end{bmatrix} \right). \quad (2.33)$$

where $\hat{\mathbf{x}}_{t+1|t} = \mathbb{E}(\mathbf{x}_{t+1} | \mathbf{y}_t, \mathbf{u}_t, \mathbf{y}_{t-1}) = \mathbb{E}(\mathbf{x}_{t+1} | \mathbf{y}_t)$ is used and similarly $\hat{\mathbf{x}}_{t|t} = \mathbb{E}(\mathbf{x}_t | \mathbf{y}_t, \mathbf{u}_t, \mathbf{y}_{t-1}) = \mathbb{E}(\mathbf{x}_t | \mathbf{y}_t)$, see (2.10). Additionally, the covariance matrix calculation uses (2.2)

$$\begin{aligned} \mathbf{P}_{t+1|t} &\equiv \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top] \\ &= \mathbb{E}[(\mathbf{x}_{t+1} - \mathbb{E}(\mathbf{x}_{t+1} | \mathbf{y}_t))(\mathbf{x}_{t+1} - \mathbb{E}(\mathbf{x}_{t+1} | \mathbf{y}_t))^\top] \end{aligned} \quad (2.34)$$

and similarly for $\mathbf{P}_{t|t}$. Additionally

$$\mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top] = \mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{F}\mathbf{x}_t + \mathbf{v}_{t+1} - \mathbf{F}\hat{\mathbf{x}}_{t|t})^\top] \quad (2.35)$$

and because \mathbf{v}_{t+1} is uncorrelated with \mathbf{x}_t and $\hat{\mathbf{x}}_{t|t}$, it yields

$$\mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top] = \mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^\top \mathbf{F}^\top] = \mathbf{P}_{t|t} \mathbf{F}^\top. \quad (2.36)$$

Then the conditional mean value is

$$\begin{aligned} \mathbb{E}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y}_t) &= \mathbb{E}(\mathbf{x}_t | \mathbf{y}_t) + (\mathbf{P}_{t|t} \mathbf{F}^\top) (\mathbf{P}_{t+1|t})^{-1} (\mathbb{E}(\mathbf{x}_{t+1} | \mathbf{y}_t) - \hat{\mathbf{x}}_{t+1|t}) \\ &= \hat{\mathbf{x}}_{t|t} + \mathbf{P}_{t|t} \mathbf{F}^\top \mathbf{P}_{t+1|t}^{-1} (\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}), \end{aligned} \quad (2.37)$$

for all t .

Defining

$$\mathbf{J}_t = \mathbf{P}_{t|t} \mathbf{F}^\top \mathbf{P}_{t+1|t}^{-1}, \quad (2.38)$$

the previous equation can be rewritten as

$$\mathbb{E}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y}_t) = \hat{\mathbf{x}}_{t|t} + \mathbf{J}_t (\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}). \quad (2.39)$$

Note that because of $\mathbb{E}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y}_T) = \mathbb{E}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y}_t)$, knowledge of \mathbf{y}_{t+j} or \mathbf{u}_{t+j} for $j > 0$ would be of no added value if the value of \mathbf{x}_{t+1} is already known. To clarify this, \mathbf{y}_{t+j} can be rewritten as

$$\mathbf{y}_{t+j} = \mathbf{A}\mathbf{u}_{t+j} + \mathbf{H}(\mathbf{F}^{j-1}\mathbf{x}_{t+1} + \mathbf{F}^{j-2}\mathbf{v}_{t+2} + \mathbf{F}^{j-3}\mathbf{v}_{t+3} + \dots + \mathbf{v}_{t+j}) + \mathbf{w}_{t+j} \quad (2.40)$$

for all $j > 0$. The error $\mathbf{x}_t - \mathbb{E}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y}_t)$ is uncorrelated with \mathbf{x}_{t+1} and \mathbf{u}_{t+j} , \mathbf{w}_{t+j} and $\mathbf{v}_{t+j}, \mathbf{v}_{t+j-1}, \dots, \mathbf{v}_{t+2}$ under the maintained assumptions. Thus, the error is uncorrelated with \mathbf{y}_{t+j} or \mathbf{u}_{t+j} for $j > 0$ resulting into

$$\mathbb{E}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y}_T) = \hat{\mathbf{x}}_{t|t} + \mathbf{J}_t(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}). \quad (2.41)$$

The first term $\hat{\mathbf{x}}_{t|t}$ indicates a particular exact linear function of \mathbf{y}_t (the coefficients of this functions are constructed from population moments and they should be viewed as deterministic constants). The term \mathbf{J}_t is also a function of population moments and is again treated as deterministic.⁴ The term $\hat{\mathbf{x}}_{t+1|t}$ is another exact linear function of \mathbf{y}_t , thus

$$\mathbb{E}(\mathbf{x}_t | \mathbf{y}_T) = \hat{\mathbf{x}}_{t|t} + \mathbf{J}_t (\mathbb{E}\hat{\mathbf{x}}_{t+1} | \mathbf{y}_T - \hat{\mathbf{x}}_{t+1|t}), \quad (2.42)$$

or equivalently

$$\hat{\mathbf{x}}_{t|T} = \hat{\mathbf{x}}_{t|t} + \mathbf{J}_t(\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}), \quad (2.43)$$

for $t = T - 1, \dots, 1$ and $\hat{\mathbf{x}}_{T|T} = \hat{\mathbf{x}}_T$.

Calculation of the MSE associated with the smoothed estimate, starting from (2.43) and subtracting both sides from \mathbf{x}_t , yields

$$\mathbf{x}_t - \hat{\mathbf{x}}_{t|T} = \mathbf{x}_t - \hat{\mathbf{x}}_{t|t} - \mathbf{J}_t \hat{\mathbf{x}}_{t+1|T} + \mathbf{J}_t \hat{\mathbf{x}}_{t+1|t} \quad (2.44)$$

$$\mathbf{x}_t - \hat{\mathbf{x}}_{t|T} + \mathbf{J}_t \hat{\mathbf{x}}_{t+1|T} = \mathbf{x}_t - \hat{\mathbf{x}}_{t|t} + \mathbf{J}_t \hat{\mathbf{x}}_{t+1|t}. \quad (2.45)$$

Multiplying the equation by its transpose and taking expectations yield

$$\begin{aligned} \mathbb{E} [(\mathbf{x}_t - \hat{\mathbf{x}}_{t|T})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|T})^\top] + \mathbf{J}_t \mathbb{E} [\hat{\mathbf{x}}_{t+1|T} \hat{\mathbf{x}}_{t+1|T}^\top] \mathbf{J}_t^\top = \\ \mathbb{E} [(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^\top] + \mathbf{J}_t [\hat{\mathbf{x}}_{t+1|t} \hat{\mathbf{x}}_{t+1|t}^\top] \mathbf{J}_t^\top, \end{aligned} \quad (2.46)$$

which states

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t [-\mathbb{E}(\hat{\mathbf{x}}_{t+1|T} \hat{\mathbf{x}}_{t+1|T}^\top) + \mathbb{E}(\hat{\mathbf{x}}_{t+1|t} \hat{\mathbf{x}}_{t+1|t}^\top)] \mathbf{J}_t^\top. \quad (2.47)$$

The cross-product terms disappeared because $\hat{\mathbf{x}}_{t+1|T}$ is a linear function of \mathbf{y}_T and is uncorrelated with the projection error $\mathbf{x}_t - \hat{\mathbf{x}}_{t|T}$. Similarly, $\hat{\mathbf{x}}_{t+1|t}$ is uncorrelated with $\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}$.

Notice the following relations

$$\begin{aligned} \mathbb{E} [\mathbf{x}_{t+1} \hat{\mathbf{x}}_{t+1|T}^\top] &= \mathbb{E} [(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|T}^\top + \hat{\mathbf{x}}_{t+1|T}^\top) \hat{\mathbf{x}}_{t+1|T}^\top] \\ &= \mathbb{E} [(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|T}^\top) \hat{\mathbf{x}}_{t+1|T}^\top] + \mathbb{E} [\hat{\mathbf{x}}_{t+1|T}^\top \hat{\mathbf{x}}_{t+1|T}^\top] \\ &= \mathbb{E} [\hat{\mathbf{x}}_{t+1|T}^\top \hat{\mathbf{x}}_{t+1|T}^\top], \end{aligned} \quad (2.48)$$

because the projection error $(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|T}^\top)$ is uncorrelated with $\hat{\mathbf{x}}_{t+1|T}^\top$. Similarly

$$\mathbb{E}(\mathbf{x}_{t+1} \hat{\mathbf{x}}_{t+1|t}^\top) = \mathbb{E}(\hat{\mathbf{x}}_{t+1|t}^\top \hat{\mathbf{x}}_{t+1|t}^\top). \quad (2.49)$$

⁴A theory about linear projection is solely used for derivation of the smoothing equation in [32]. The previously used approach employing properties of the multivariate normal distribution can be an alternative way of the derivation proposed by the author of this thesis.

Let's recalculate the bracketed term in (2.47) using the previous two relations

$$\begin{aligned}
& -\mathbb{E}(\hat{\mathbf{x}}_{t+1|T} \hat{\mathbf{x}}_{t+1|T}^\top) + \mathbb{E}(\hat{\mathbf{x}}_{t+1|t} \hat{\mathbf{x}}_{t+1|t}^\top) \\
&= [\mathbb{E}(\mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top) - \mathbb{E}(\hat{\mathbf{x}}_{t+1|T} \hat{\mathbf{x}}_{t+1|T}^\top)] - [\mathbb{E}(\mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top) - \mathbb{E}(\hat{\mathbf{x}}_{t+1|t} \hat{\mathbf{x}}_{t+1|t}^\top)] \\
&= \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|T})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|T})^\top] - \mathbb{E}[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})^\top] \\
&= \mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}. \tag{2.50}
\end{aligned}$$

Substituting the previous result back into (2.47) establishes that the smoothed estimate $\hat{\mathbf{x}}_{t|T}$ has the MSE given by

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{J}_t^\top \tag{2.51}$$

for $t = T - 1, \dots, 1$ and $\mathbf{P}_{T|T} = \mathbf{P}_T$.

From the previous is clear that the smoothed algorithm requires that sequences $\{\hat{\mathbf{x}}_{t|t}\}_{t=1}^T$, $\{\hat{\mathbf{x}}_{t+1|t}\}_{t=1}^T$, $\{\mathbf{P}_{t|t}\}_{t=1}^T$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^T$ are stored for all t so that they can be combined with $\hat{\mathbf{x}}_{t+1|T}$ and $\mathbf{P}_{t+1|T}$. If $\mathbf{P}_{t+1|T}$ is singular for some t , it may be replaced by a generalized inverse as it is suggested in [45].

2.1.4 The Kalman Filter: Set of Equations

To conclude the derivation of the Kalman filter equations for prediction and updating, the key equations are reproduced here again. These are (2.13), (2.14), (2.16) and (2.19)

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \tag{2.52}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_{t|t-1} \tag{2.53}$$

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t} \tag{2.54}$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^\top + \mathbf{Q}. \tag{2.55}$$

Equation (2.54) can be equivalently replaced by (2.17)

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \tag{2.56}$$

and (2.55) can be substituted either by (2.20) or by (2.23)

$$\mathbf{P}_{t+1|t} = \mathbf{F}[\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_{t|t-1}]\mathbf{F}^\top + \mathbf{Q} \tag{2.57}$$

$$\mathbf{P}_{t+1|t} = (\mathbf{F} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1}(\mathbf{F}^\top - \mathbf{H}^\top\mathbf{K}_t^\top) + \mathbf{K}_t\mathbf{R}\mathbf{K}_t^\top + \mathbf{Q}. \tag{2.58}$$

where the (Kalman) gain matrix, defined as (2.18), is

$$\mathbf{K}_t \equiv \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}. \tag{2.59}$$

Sometimes, see e.g. [67], [72], or [89], the Kalman gain matrix (2.59) is defined alternatively as

$$\dot{\mathbf{K}}_t \equiv \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}, \tag{2.60}$$

which means that $\dot{\mathbf{K}}_t = \mathbf{F}\mathbf{K}_t$ for all t . This also implies that (2.52) and (2.53) can be rewritten

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \dot{\mathbf{K}}_t(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \tag{2.61}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \dot{\mathbf{K}}_t\mathbf{H}\mathbf{P}_{t|t-1} = (\mathbf{I} - \dot{\mathbf{K}}_t\mathbf{H})\mathbf{P}_{t|t-1}. \tag{2.62}$$

This definition of $\dot{\mathbf{K}}_t$ expresses an idea that the gain matrix is chosen to minimize the MSE of the update $\mathbf{P}_{t|t}$, see [12] and [88] or [24]. This minimization can be accomplished by substituting equation (2.61) in a forecasting estimate error ($\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}$), substituting this into definition of $\mathbf{P}_{t|t}$, taking the derivative of the trace of the result with respect to \mathbf{K}_t , setting that result equal to zero and then solving for \mathbf{K}_t , resulting in (2.60).⁵ Looking at (2.60), it is clear that as the measurement error covariance \mathbf{R} approaches zero, the gain $\dot{\mathbf{K}}_t$ weights the residual more heavily. Specifically,

$$\lim_{\mathbf{R}_t \rightarrow \mathbf{0}} \dot{\mathbf{K}}_t = \mathbf{H}^{-1}. \quad (2.63)$$

Contrary, as the estimate error covariance $\mathbf{P}_{t|t-1}$ approaches zero, the gain $\dot{\mathbf{K}}_t$ weights the residual less heavily. Specifically,

$$\lim_{\mathbf{P}_{t|t-1} \rightarrow \mathbf{0}} \dot{\mathbf{K}}_t = \mathbf{0}. \quad (2.64)$$

From the computational point, using the Kalman filter equations written in the previous form is not necessarily the best way to proceed. An alternative algorithm is the information filter, see [4] and [34]. Rather than yielding a set of recursion for the MSE matrix \mathbf{P}_t , the information filter gives a set of its inverse \mathbf{P}_t^{-1} , which is known as the information matrix. The filter gives recursions for the vector $\mathbf{P}_t^{-1}\hat{\mathbf{x}}_t$. The information filter is quite convenient when the initial covariance matrix \mathbf{P}_0 is infinite, since then $\mathbf{P}_0^{-1} = 0$. It also appears to become attractive when a number of series is significantly greater than the dimension of the state. This is because it is not required to calculate the inverse matrix in (2.52), (2.53), (2.57) and (2.59) respectively.⁶

Another option is to use the square filter which operates on matrix $\dot{\mathbf{P}}_t$ such that $\mathbf{P}_t = \dot{\mathbf{P}}_t \dot{\mathbf{P}}_t$. By doing so, the problem that the Kalman filter may failure due to \mathbf{P}_t not being non-negative definite is avoided. This algorithm is also more numerically stable than the previous ones, but it requires higher computational burden. For further discussion see [34].

Once the system of equation for the Kalman filtering is introduced, it is easy to forecast \mathbf{y}_{t+1} and calculate the associated MSE

$$\hat{\mathbf{y}}_{t+1|t} \equiv \mathbb{E}(\mathbf{y}_{t+1} | \mathbf{u}_{t+1}, \mathbf{xy}_t) = \mathbf{A}\mathbf{u}_{t+1} + \mathbf{H}\hat{\mathbf{x}}_{t+1|t}, \quad (2.65)$$

$$\mathbb{E}[(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})^\top] = \mathbf{H}\mathbf{P}_{t+1|t}\mathbf{H}^\top + \mathbf{R}. \quad (2.66)$$

For the MSE calculation, equation (2.9) was used.

The smoothing procedure by the Kalman filter is composed by these equations (2.43), (2.51) and (2.38)

$$\hat{\mathbf{x}}_{t|T} = \hat{\mathbf{x}}_{t|t} + \mathbf{J}_t(\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}) \quad (2.67)$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{J}_t^\top, \quad (2.68)$$

⁵The original way of thinking about the weighting by \mathbf{K}_t , used for derivation of the gain, is that as the measurement error covariance \mathbf{R} approaches zero, the actual measurement \mathbf{y}_t is more and more reliable, while the predicted measurement $\hat{\mathbf{y}}_{t|t-1}$ is trusted less and less. On the other hand, as $\mathbf{P}_{t|t-1}$ approaches zero the actual measurement \mathbf{y}_t is trusted less and less, while the predicted measurement $\hat{\mathbf{y}}_{t|t-1}$ is trusted more and more.

⁶It is generally assumed that the inverse matrix exists. This can be replaced by a pseudo-inverse.

where

$$\mathbf{J}_t = \mathbf{P}_{t|t} \mathbf{F}^\top \mathbf{P}_{t+1|t}^{-1} \quad (2.69)$$

for $t = T - 1, \dots, 1$.

2.2 Maximum Likelihood Estimation

So far it was used an assumption that the coefficients in the matrices \mathbf{F} , \mathbf{Q} , \mathbf{A} , \mathbf{H} and \mathbf{R} are known. In case the values of the matrices are unknown, the Kalman filter can be used to evaluate the likelihood function. It can be proved that the forecast of $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{y}}_{t|t-1}$ are optimal within the set of forecasts that are linear in $(\mathbf{u}_t, \mathbf{xy}_{t-1})$, see [72]. If the initial state \mathbf{x}_1 and the innovations $\{\mathbf{v}_t, \mathbf{w}_t\}_{t=1}^T$ are multivariate Gaussian, then a stronger claim that the forecasts $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{y}}_{t|t-1}$ calculated by the Kalman filter are optimal among any functions of $(\mathbf{u}_t, \mathbf{xy}_{t-1})$, see [72]. Moreover, if \mathbf{x}_1 and $\{\mathbf{v}_t, \mathbf{w}_t\}_{t=1}^T$ are Gaussian, then the distribution of \mathbf{y}_t conditional on $(\mathbf{u}_t, \mathbf{xy}_{t-1})$ is Gaussian with mean given by (2.65) and variance given by (2.66), so

$$(\mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1}) \sim \mathcal{N}((\mathbf{A}\mathbf{u}_t + \mathbf{H}\hat{\mathbf{x}}_{t|t-1}), (\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})), \quad (2.70)$$

which is

$$p(\mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1}) = (2\pi)^{-\frac{n}{2}} |\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1})^\top (\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1} (\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1})\right\}, \quad (2.71)$$

for $t = 1, 2, \dots, T$

From the previous equation, it is possible to construct the sample log likelihood in the standard way: $\sum_{t=1}^T \log p(\mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1})$.

2.2.1 Maximum Likelihood Estimation of Parameters

When the values of matrices \mathbf{F} , \mathbf{Q} , \mathbf{A} , \mathbf{H} and \mathbf{R} are not known, it is possible to proceed as follows. Collect the unknown elements of these matrices in a vector θ . Make an arbitrary initial guess as to a value of θ , denoted $\theta^{(0)}$, and calculate the sequences $\{\hat{\mathbf{x}}_{t|t-1}(\theta^{(0)})\}_{t=1}^T$ and $\{\mathbf{P}_{t|t-1}(\theta^{(0)})\}_{t=1}^T$ that result from this value in (2.17) and (2.20). Recall from (2.12) that if data are generated from the state-space model (1.13) – (1.16) with this value of θ , then

$$(\mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1}, \theta^{(0)}) \sim \mathcal{N}(\mu_t(\theta^{(0)}), \Sigma_t(\theta^{(0)})), \quad (2.72)$$

where

$$\mu_t(\theta^{(0)}) = [\mathbf{A}(\theta^{(0)})]\mathbf{u}_t + [\mathbf{H}(\theta^{(0)})]\hat{\mathbf{x}}_{t|t-1}(\theta^{(0)}), \quad (2.73)$$

$$\Sigma_t(\theta^{(0)}) = [\mathbf{H}(\theta^{(0)})][\mathbf{P}_{t|t-1}(\theta^{(0)})][\mathbf{H}(\theta^{(0)})]^\top + [\mathbf{R}(\theta^{(0)})]. \quad (2.74)$$

The value of the log likelihood is then

$$\begin{aligned} \sum_{t=1}^T \log p(\mathbf{y}_t | \mathbf{u}_t, \mathbf{xy}_{t-1}, \theta^{(0)}) &= -\frac{nT}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log |\Sigma_t(\theta^{(0)})| - \\ &\quad - \frac{1}{2} \sum_{t=1}^T \{[\mathbf{y}_t - \mu_t(\theta^{(0)})]^\top [\Sigma_t(\theta^{(0)})]^{-1} [\mathbf{y}_t - \mu_t(\theta^{(0)})]\}, \end{aligned} \quad (2.75)$$

which reflects how likely it would have been to have observed data if $\theta^{(0)}$ were the true value for θ .⁷

The maximum likelihood estimation procedure supposes that the model is identified. This assumes that a change in any of the parameters would imply a different probability distribution for $\{\mathbf{y}_t\}_{t=1}^T$.

Although the state-space representation gives a convenient way to calculate the exact likelihood function, a problem of identification can occur. In the absence of restrictions on \mathbf{F} , \mathbf{Q} , \mathbf{A} , \mathbf{H} or \mathbf{R} , the parameters are unidentified - more that one set of the parameters can give the identical value of the likelihood function. Following [69], the model is not either globally or locally identified.

2.2.2 Asymptotic Properties

The maximum likelihood estimate $\hat{\theta}$ based on the sample of size T is consistent and asymptotically normal if the following conditions are satisfied, see [13]

- the model is identified,
- eigenvalues of \mathbf{F} are all inside the unit circle,
- the exogenous variable \mathbf{u}_t (i.e. apart from a constant term) behave asymptotically like a full rank linearly non-deterministic covariance-stationary process, and
- the true value of θ does not fall on the boundary of the allowable parameter space.

2.3 The Kalman Filter Alterations

The derivation of the linear Kalman filter assumed no correlation between \mathbf{v}_t and \mathbf{w}_t for all t . However, this is straightforward to generalize, see [4]. Predetermined or exogenous variables can also be added to the state equation with few adjustments, see e.g. [70] or [90]. Moreover, the Kalman filter is a convenient algorithm for handling missing observations. If \mathbf{y}_t , is unobserved at some time t , one can simply skip the updating equations (2.52) and (2.53) for this date and replace them with $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1}$ and $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1}$, see e.g. [46]. Modifications of the Kalman filtering and smoothing algorithms to allow for singular $\mathbf{P}_{t|t}$ are described in [26].

In the following two subsections, an assumption of normality of disturbances and time varying parameters as additional alterations of the Kalman filter are discussed in more details.

⁷In case of a notation simplification and $\theta^{(0)}$ parameter omission, it is possible to write

$$\sum_{t=1}^T \log p(\mathbf{y}_t) = -\frac{Tn}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log |\boldsymbol{\Sigma}_t| - \frac{1}{2} \sum_{t=1}^T \nu_t^\top \boldsymbol{\Sigma}_t^{-1} \nu_t,$$

where $\nu_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} = \mathbf{y}_t - \mathbb{E}\mathbf{y}_{t|t-1} = \mathbf{y}_t - (\mathbf{A}\mathbf{u}_t + \mathbf{H}\mathbf{x}_{t|t-1})$ for $t = 1, \dots, T$ using the state equation (1.15). The equation is sometimes known as the prediction error decomposition form of the likelihood, see [34].

2.3.1 The Kalman Filter with Non-Normal Disturbances

In case that the disturbances \mathbf{v}_t and \mathbf{w}_t are not normally distributed, the Kalman filter can be calculated in the same way.

Under an assumption of normality, $\hat{\mathbf{x}}_{t|t-1}$ is the function of \mathbf{xy}_{t-1} that minimizes the MSE (see the definition of $\mathbf{P}_{t+1|t}$ by equation (2.2))

$$\mathbb{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\top] \quad (2.76)$$

in the sense that any other forecast has a mean squared error matrix that differs from that of $\hat{\mathbf{x}}_{t|t-1}$ by a positive semidefinite matrix. This optimal forecast turned out to be a constant plus a linear function of \mathbf{xy}_{t-1} . The minimum value achieved for (2.76) was denoted $\mathbf{P}_{t|t-1}$.

If \mathbf{v}_t and \mathbf{w}_t are not normal, one can pose a related problem of choosing $\hat{\mathbf{x}}_{t|t-1}$ to be a constant plus a linear function of \mathbf{xy}_{t-1} that minimizes (2.76). The solution to this problem turns out to be given by the Kalman filter iteration (2.56) and its unconditional mean squared error is still given by (2.56). Similarly, when the disturbances are not normal, expression (2.31) can be interpreted as the linear projection of \mathbf{y}_{t+m} on \mathbf{xy}_{t-1} , and a constant, with (2.32) its unconditional mean squared error. Thus, while the Kalman filter forecasts need no longer be optimal for systems that are not normal, no other forecast based on a linear function of \mathbf{xy}_{t-1} will have a smaller mean squared error. See a discussion in [4] for more details.

Also function (2.75) is formed in the same way and maximized with respect to θ even for non-normal systems. This quasi-maximum likelihood estimations still yield consistent and asymptotically normal estimates of the elements of \mathbf{F} , \mathbf{Q} , \mathbf{A} , \mathbf{H} and \mathbf{R} , see [87].

2.3.2 Time-Varying Coefficients

The analysis above treated the coefficients of the matrices \mathbf{F} , \mathbf{Q} , \mathbf{A} , \mathbf{H} and \mathbf{R} as known constants. One of a possible generalization is obtained if these are known functions of \mathbf{u}_t

$$\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{u}_t)\mathbf{x}_t + \mathbf{v}_{t+1}, \quad (2.77)$$

$$\mathbb{E}(\mathbf{v}_{t+1}\mathbf{v}_{t+1}^\top) = \mathbf{Q}(\mathbf{u}_t), \quad (2.78)$$

$$\mathbf{y}_t = \mathbf{a}(\mathbf{u}_t) + \mathbf{H}(\mathbf{u}_t)\mathbf{x}_t + \mathbf{w}_t, \quad (2.79)$$

$$\mathbb{E}(\mathbf{w}_t\mathbf{w}_t^\top) = \mathbf{R}(\mathbf{u}_t), \quad (2.80)$$

where $\mathbf{F}(\mathbf{u}_t)$, $\mathbf{Q}(\mathbf{u}_t)$, $\mathbf{H}(\mathbf{u}_t)$ and $\mathbf{R}(\mathbf{u}_t)$ denote matrix-valued functions of \mathbf{u}_t , and $\mathbf{a}(\mathbf{u}_t)$ is an $(n \times 1)$ vector-valued function of \mathbf{u}_t . As before, it is assumed that \mathbf{u}_t provides no information about \mathbf{x}_τ or \mathbf{w}_τ for any τ beyond that contained in \mathbf{xy}_{t-1} .

Even if \mathbf{v}_t and \mathbf{w}_t are normal with \mathbf{u}_t stochastic, the unconditional distributions of \mathbf{x}_t and \mathbf{y}_t are no longer normal. However, the system is conditionally normal in the following way. Suppose that the distribution of \mathbf{x}_t conditional on \mathbf{u}_t and \mathbf{xy}_{t-1} is $\mathcal{N}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$. Then \mathbf{x}_t conditional on \mathbf{u}_t and \mathbf{xy}_{t-1} has the same distribution. Moreover, conditional on \mathbf{u}_t , all of the matrices can be treated as

deterministic. Hence the derivation of the Kalman filter goes through essentially as before, with the recursions (2.56) and (2.57) replaced by

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}(\mathbf{u}_t)\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t [\mathbf{y}_t - \mathbf{a}(\mathbf{u}_t) - \mathbf{H}(\mathbf{u}_t)\hat{\mathbf{x}}_{t|t-1}] \quad (2.81)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}(\mathbf{u}_t) \left\{ \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{u}_t)^\top [\mathbf{H}(\mathbf{u}_t)\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{u}_t)^\top + \mathbf{R}(\mathbf{u}_t)]^{-1} \right. \\ \left. [\mathbf{H}(\mathbf{u}_t)]^\top \mathbf{P}_{t|t-1} \right\} \mathbf{F}(\mathbf{u}_t)^\top + \mathbf{Q}(\mathbf{u}_t) \quad (2.82)$$

$$\mathbf{K}_t \equiv \mathbf{F}(\mathbf{u}_t)\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{u}_t)^\top [\mathbf{H}(\mathbf{u}_t)\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{u}_t)^\top + \mathbf{R}(\mathbf{u}_t)]^{-1} \quad (2.83)$$

It is worth noting three elements of the earlier discussion that was changed by the time-varying parameter matrices

- the distribution calculated for the initial state in (2.4) and (2.5) is only valid if $\mathbf{F}(\mathbf{u}_t)$ and $\mathbf{Q}(\mathbf{u}_t)$ are fixed matrices,
- m -period-ahead forecasts of \mathbf{y}_{t+m} or \mathbf{x}_{t+m} for $m > 1$ are no longer simple to calculate when $\mathbf{F}(\mathbf{u}_t)$, $\mathbf{H}(\mathbf{u}_t)$, or $\mathbf{A}(\mathbf{u}_t)$ vary stochastically, and
- if \mathbf{v}_t , and \mathbf{w}_t are not normal, then the one-period-ahead forecasts of $\hat{\mathbf{x}}_{t+1|t}$ and $\hat{\mathbf{y}}_{t+1|t}$ are no longer interpreted as linear projections, since (2.81) is non-linear in \mathbf{u}_t .

The time-varying coefficient model with discrete-valued state variables can be useful way of dealing with changes occurring in economic policies. These changes often take a form of dramatic and discrete events. It is thus of interest to consider time-series models in which coefficients change only occasionally as a result of such changes (especially in case of changes in policy regimes). The discrete analog to the state equation is based on an assumption that the probability distribution of the state vector depends on past events only through its lagged value. When this probabilities does not depend on the previous state, the system is the Markov-switching model (originally developed by [29] and further elaborated by others).⁸

2.4 Linear Constraints

Generally, the constraints can be of two basic types - equality and inequality. The thesis copes only with the inequality ones.

There are two ways how to incorporate state inequality constraints, see [73].

Hard (Strong) Constraints. The first method incorporated hard constraints to maintain the state variable estimates within a user-defined envelope. The constraints represent some relations between state variables which are known exactly and hence are used to describe the relation.

In case of the Kalman filter and hard inequality constraints, the resultant filter is a combination of a standard Kalman filter and a quadratic programming problem.

Soft (Weak) Constraints. The second method incorporates soft constraints to ensure that the state variable estimates vary slowly with time. The soft constraints are constraints that are required to be approximately rather than exactly

⁸In the Markov-switching model, the matrix \mathbf{F} in (2.81) represents a Markov chain matrix with probabilities of transitions among states.

satisfied. They are used in case that the constraints are not rigorous or the constraint function has some uncertainty.

We impose inequality constraints to our system in the following way

$$\mathbf{D}\mathbf{x}_t \leq \mathbf{d} \quad (2.84)$$

where \mathbf{D} is a known matrix and \mathbf{d} is a known vector. In this chapter several ways how to modify the Kalman filter to incorporate these constraints are introduced.

The Kalman filter with constraints in a form of (2.84) is constructed by directly projecting the unconstrained state estimate $\hat{\mathbf{x}}_t$ onto the constrained surface $\mathbf{S} = \{\mathbf{x}_t : \mathbf{D}\mathbf{x}_t \leq \mathbf{d}\}$, $\mathbf{S} \subseteq \mathbb{R}^r$. In case of the Kalman gain, this constraint surface is labeled as \mathbf{SS} ($\mathbf{SS} \subseteq \mathbb{R}^{r \times n}$), see later.

The constraint is to be applied using the architecture shown in Figure 2.1. The filter is initialized, predicted, and updated with the measurement to give an unconstrained estimate of the state vector with the covariance matrix. The constraint is applied, and the resulting estimate obeys the required constraint. This form is consistent with Alouani's suggestion: the constraint is only applied to the updated estimate and is thus likely to be most accurate, see [2].

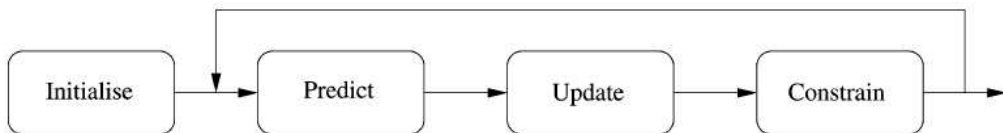


Figure 2.1: Constrained Filter Approach (Source: [36])

Inequality constraints are inherently more complicated than equality constraints, but standard quadratic programming results can be used to solve the Kalman filter problem with inequality constraints.

In case of inequality constraints, an active set method can be applied. For each constraint it is tested if the constraint is satisfied (then it is considered as an equality constraint, $\mathbf{D}\mathbf{x}_t = \mathbf{d}$) or the inequality is not satisfied. The same approach is used in case of soft constraints. Thus all the inequality constraints are transformed to the equality ones, see [76].

Generally, it is possible to suppose that at each time step $t = 1, \dots, T$, \mathbf{x}_t is subject to the following linear inequality constraint $\mathbf{D}_t\mathbf{x}_t \leq \mathbf{d}_t$. Furthermore, \mathbf{D}_t is supposed to be a full rank matrix (if it is not the case, it means that there are some redundant constraints, which can be eliminated).

There are two possibilities of using the constraints: either the updated state estimate will satisfy the constraint at each iteration, as below

$$\mathbf{D}\hat{\mathbf{x}}_{t|t} \leq \mathbf{d}, \quad (2.85)$$

or the state prediction to be constrained, which would allow a better forecast for the system

$$\mathbf{D}\hat{\mathbf{x}}_{t+1|t} \leq \mathbf{d}. \quad (2.86)$$

In the following subsections, we will mainly discuss constraining the updated state estimate. The basic principle for the forecast state estimate holds as well.

A weaker but more general constraint may exist on the expected value of the state, see [75]

$$\mathbf{D}\mathbb{E}\mathbf{x}_t \leq \mathbf{d}_t. \quad (2.87)$$

The incorporation of state variable constraints commonly increases the computational effort of the filter but significantly improves its estimation accuracy. Using an appropriate estimation method also depends on a particular application.

2.4.1 Estimate Projection

One possible approach to the constrained filtering is to project the unconstrained estimate $\hat{\mathbf{x}}_{t|t}$ of the Kalman filter onto the constraint surface, see [75]. The estimate projection approach has an advantage that it can be applied to both equality and inequality constraints. This problem is a quadratic programming problem, see e.g. [27].

Generally, we want to estimate the state \mathbf{x}_t satisfying the linear constraint. We keep a notation that $\hat{\mathbf{x}}_t$ is the state estimated at time t by an unconstrained estimator (i.e. an estimator which does not take into account any constraint). Then $\tilde{\mathbf{x}}_t$ is an estimate provided by same estimator, but which take into account a linear constraint. The principle of the projection approach is illustrated in Figure 2.2. For further discussion see [76].

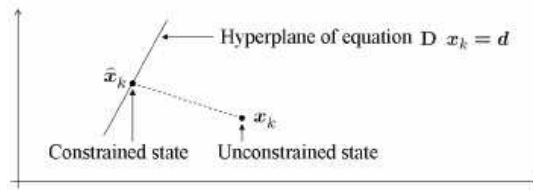


Figure 2.2: Principle of the Projection Approach (Source: [76] and own adjustments)

The Projection Method. At each time step of the constrained Kalman filter, a quadratic programming problem is solved to obtain the constrained state estimate. Here a brief tour through a transformation of the original problem to a form of the quadratic dynamic problem is presented first. It is based on [74].

From [4] is clear that the Kalman filter estimates of $\hat{\mathbf{x}}_{t|t}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$ are jointly Gaussian, in which case $\hat{\mathbf{x}}_{t|t}$ is conditionally Gaussian with given $\{\mathbf{y}_1, \dots, \mathbf{y}_t\}$. The PDF of \mathbf{x}_t given $\{\mathbf{y}_1, \dots, \mathbf{y}_t\}$ is

$$p(\mathbf{x}_t | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t) = \frac{1}{(2\pi)^{r/2} |\mathbf{P}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^\top \mathbf{P}^{-1} (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}) \right\}, \quad (2.88)$$

where r is a dimension of \mathbf{x}_t and the covariance of the conditional distribution is $\mathbf{P} = \mathbf{P}_{t|t}$ for all t , see equations (2.13) and (2.14) together with the proper footnote in Subsection 2.1.2.

The constrained Kalman filter can be derived using the maximum probability method by finding an estimate \mathbf{x} such that the conditional PDF is maximized and \mathbf{x} satisfies the required constraints (without loss of generality an inequality

constraint is used). Maximizing PDF is the same as maximizing its natural logarithm. So the problem can be given by

$$\max \log p(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t) \Rightarrow \min_{\mathbf{x} \in \mathbf{S}} (\mathbf{x} - \hat{\mathbf{x}}_{t|t})^\top \mathbf{P}^{-1} (\mathbf{x} - \hat{\mathbf{x}}_{t|t}) \quad (2.89)$$

such that

$$\mathbf{D}\mathbf{x} \leq \mathbf{d}. \quad (2.90)$$

The problem can be rewritten as

$$\min_{\mathbf{x} \in \mathbf{S}} (\mathbf{x}^\top \mathbf{P}^{-1} \mathbf{x} - 2\hat{\mathbf{x}}_{t|t}^\top \mathbf{P}^{-1} \mathbf{x}) \quad \text{such that} \quad \mathbf{D}\mathbf{x} \leq \mathbf{d}. \quad (2.91)$$

From the previous formulation is clear that that this problem statement depends on the conditional Gaussian nature of $\hat{\mathbf{x}}_{t|t}$, which in turn depends on the Gaussian nature of \mathbf{x}_0 , $\{\mathbf{v}_1, \dots, \mathbf{v}_t\}$ and $\{\mathbf{w}_1, \dots, \mathbf{w}_t\}$.⁹

The constrained Kalman filtering problem can also be specified directly by projecting the unconstrained state estimate $\hat{\mathbf{x}}_{t|t}$ onto the constraint surface. The problem is

$$\min_{\mathbf{x} \in \mathbf{S}} (\mathbf{x} - \hat{\mathbf{x}}_{t|t})^\top \mathbf{W} (\mathbf{x} - \hat{\mathbf{x}}_{t|t}) \quad \text{such that} \quad \mathbf{D}\mathbf{x} \leq \mathbf{d}, \quad (2.92)$$

where \mathbf{W} is an symmetric positive definite weighting matrix. The problem can be again equivalently rewritten as

$$\min_{\mathbf{x} \in \mathbf{S}} (\mathbf{x}^\top \mathbf{W} \mathbf{x} - 2\hat{\mathbf{x}}_{t|t}^\top \mathbf{W} \mathbf{x}) \quad \text{such that} \quad \mathbf{D}\mathbf{x} \leq \mathbf{d}. \quad (2.93)$$

The constrained estimation problems derived by the maximum probability method and the mean square method can be obtained from the previous equation by setting $\mathbf{W} = \mathbf{P}^{-1}$ and $\mathbf{W} = \mathbf{I}$ respectively. Note also that this derivation of the constrained estimation problem does not depend on the conditional Gaussian nature of $\hat{\mathbf{x}}_{t|t}$.

Estimate Projection with Inequality Constraints. If the constraint is in a form $\mathbf{D}\hat{\mathbf{x}}_t \leq \mathbf{d}$, then a constrained estimate can be obtained by the problem of the estimate projection with equality constraints. It solves the problem specified as (2.92), which is a quadratic programming problem. Various approaches can be used to solve these problems. One of a possible methods are interior point approaches and active set methods, see e.g. [11].

An active set method uses the fact that only those constraints, that are active at the solution of the problem, are significant in the optimality conditions. Suppose that there is totally s inequality constraints, and only q of the s inequality constraints are active at the solution of the quadratic programming system. The method consists in testing at each time step and checks the inequality constraints. For each inequality only two scenarii can occur: either the inequality is satisfied

⁹ The constrained Kalman filtering problem can also be derived by using a mean square minimization method. The conditional mean square error is minimized subject to the state constraints

$$\min_{\mathbf{x} \in \mathbf{S}} \mathbb{E}(\|\hat{\mathbf{x}}_{t|t} - \mathbf{x}\|^2 | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t) \quad \text{such that} \quad \mathbf{D}\mathbf{x} \leq \mathbf{d}.$$

The first order condition necessary for a minimum, after some algebra (see [75]), is the same as equation (2.91).

and so do not have to be taken in account, or the inequality is not satisfied and it is applied to a boundary. The active set is made up of those inequality constraints with non-zero Lagrange multipliers.

Denote by $\tilde{\mathbf{D}}$ the q rows of \mathbf{D} that correspond to the active constraints, and denote by $\tilde{\mathbf{d}}$ the corresponding q components of \mathbf{d} that are linked to the active constraints. If the set of active constraints is known then the solution of this problem is also a solution of the equality constrained problem

$$\tilde{\mathbf{x}}_{t|t} = \underset{\mathbf{x} \in \mathbf{S}}{\operatorname{argmin}} (\mathbf{x} - \hat{\mathbf{x}}_{t|t})^\top \mathbf{W} (\mathbf{x} - \hat{\mathbf{x}}_{t|t}) \quad \text{such that} \quad \tilde{\mathbf{D}}\mathbf{x} = \tilde{\mathbf{d}}. \quad (2.94)$$

The inequality constrained problem is thus equivalent to the problem with equality constraints, which is introduced in the following paragraph. All of the properties of the equality constrained state estimate are applied to the inequality constrained state estimate.

Estimate Projection with Equality Constraints. The constrained estimate can be written as

$$\tilde{\mathbf{x}}_{t|t} = \underset{\mathbf{x} \in \mathbf{S}}{\operatorname{argmin}} (\mathbf{x} - \hat{\mathbf{x}}_{t|t})^\top \mathbf{W} (\mathbf{x} - \hat{\mathbf{x}}_{t|t}) \quad \text{such that} \quad \mathbf{D}\mathbf{x} = \mathbf{d}, \quad (2.95)$$

where \mathbf{W} is a symmetric positive-definite weighting matrix.¹⁰ The solution to this problem is

$$\tilde{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t} - \mathbf{W}^{-1}\mathbf{D}^\top (\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^\top)^{-1} (\mathbf{D}\hat{\mathbf{x}}_{t|t} - \mathbf{d}) \quad (2.96)$$

$$= \hat{\mathbf{x}}_{t|t} - \mathbf{C} (\mathbf{D}\hat{\mathbf{x}}_{t|t} - \mathbf{d}), \quad (2.97)$$

where $\mathbf{C} = \mathbf{W}^{-1}\mathbf{D}^\top (\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^\top)^{-1}$. Using a reduced form for $\mathbf{x}_t - \tilde{\mathbf{x}}_{t|t}$ and the

¹⁰ Generally, the problem can be specified

$$\tilde{\mathbf{x}}_t = \underset{\tilde{\mathbf{x}} \in \mathbf{S}}{\operatorname{argmin}} (\tilde{\mathbf{x}} - \hat{\mathbf{x}}_t)^\top \mathbf{W}_t (\tilde{\mathbf{x}} - \hat{\mathbf{x}}_t) \quad \text{such that} \quad \mathbf{D}_t \tilde{\mathbf{x}}_t = \mathbf{d}_t.$$

The solution is obtained through the use of the Lagrange multiplier. The Lagrangian is

$$\mathbf{L}(t, \tilde{\mathbf{x}}, \lambda) = (\tilde{\mathbf{x}} - \hat{\mathbf{x}}_t)^\top \mathbf{W}_t (\tilde{\mathbf{x}} - \hat{\mathbf{x}}_t) + 2\lambda^\top (\mathbf{D}_t \tilde{\mathbf{x}}_t - \mathbf{d}_t).$$

The first order conditions for a minimum are

$$\begin{aligned} \frac{\partial \mathbf{L}(t, \tilde{\mathbf{x}}, \lambda)}{\partial \tilde{\mathbf{x}}} = 0 & \quad \Rightarrow \quad \mathbf{W}_t (\tilde{\mathbf{x}} - \hat{\mathbf{x}}_t) + \mathbf{D}_t^\top \lambda = 0 \\ \frac{\partial \mathbf{L}(t, \tilde{\mathbf{x}}, \lambda)}{\partial \lambda} = 0 & \quad \Rightarrow \quad \mathbf{D}_t \tilde{\mathbf{x}}_t - \mathbf{d}_t = 0. \end{aligned}$$

This gives the solution

$$\begin{aligned} \tilde{\mathbf{x}}_t &= \hat{\mathbf{x}}_t - \mathbf{W}_t^{-1}\mathbf{D}_t^\top (\mathbf{D}_t\mathbf{W}_t^{-1}\mathbf{D}_t^\top)^{-1} (\mathbf{D}_t\hat{\mathbf{x}}_t - \mathbf{d}_t) \\ \lambda &= (\mathbf{D}_t\mathbf{W}_t^{-1}\mathbf{D}_t^\top)^{-1} (\mathbf{D}_t\hat{\mathbf{x}}_t - \mathbf{d}_t). \end{aligned}$$

definition of the error covariance matrix (2.2), the following expression arrives

$$\begin{aligned}
\mathbf{x}_t - \tilde{\mathbf{x}}_{t|t} &= \mathbf{x}_t - \mathbf{x}_{t|t} + \mathbf{C} (\mathbf{D}\hat{\mathbf{x}}_{t|t} - \mathbf{d} - (\mathbf{D}\mathbf{x}_t - \mathbf{d})) \\
&= \mathbf{x}_t - \mathbf{x}_{t|t} + \mathbf{C} (\mathbf{D}\hat{\mathbf{x}}_{t|t} - \mathbf{D}\mathbf{x}_t) \\
&= -(\mathbf{I} - \mathbf{C}\mathbf{D})(\mathbf{x}_t - \mathbf{x}_{t|t})
\end{aligned} \tag{2.98}$$

$$\begin{aligned}
\tilde{\mathbf{P}}_{t|t} &= \mathbb{E}(\mathbf{x}_t - \tilde{\mathbf{x}}_{t|t})(\mathbf{x}_t - \tilde{\mathbf{x}}_{t|t})^\top \\
&= \mathbb{E}(\mathbf{I} - \mathbf{C}\mathbf{D})(\mathbf{x}_t - \mathbf{x}_{t|t})(\mathbf{x}_t - \mathbf{x}_{t|t})^\top (\mathbf{I} - \mathbf{C}\mathbf{D})^\top \\
&= (\mathbf{I} - \mathbf{C}\mathbf{D})\mathbf{P}_{t|t}(\mathbf{I} - \mathbf{C}\mathbf{D})^\top \\
&= \mathbf{P}_{t|t} - \mathbf{C}\mathbf{D}\mathbf{P}_{t|t} - \mathbf{P}_{t|t}\mathbf{D}^\top\mathbf{C}^\top + \mathbf{C}\mathbf{D}\mathbf{P}_{t|t}\mathbf{P}_{t|t}\mathbf{D}^\top\mathbf{C}^\top \\
&= \mathbf{P}_{t|t} - \mathbf{C}\mathbf{D}\mathbf{P}_{t|t} \\
&= (\mathbf{I} - \mathbf{C}\mathbf{D})\mathbf{P}_{t|t}.
\end{aligned} \tag{2.99}$$

If the process and measurement noises are Gaussian and $\mathbf{W} = \mathbf{P}_{t|t}^{-1}$ is set, the maximum probability estimate of the state subject to state constraints is obtained. Moreover, the constrained estimated state has the following properties, see [75] and [76]

- it is unbiased for any symmetric positive definite weighting matrix \mathbf{W} , for all $t = 1, \dots, T$: $\mathbb{E}\mathbf{x}_t = \mathbb{E}\tilde{\mathbf{x}}_t$,
- if $\mathbf{W} = \mathbf{P}_{t|t}^{-1}$, then it has an error covariance that is less than or equal to that of the unconstrained state estimate: $cov(\mathbf{x}_t - \tilde{\mathbf{x}}_t) \leq cov(\mathbf{x}_t - \hat{\mathbf{x}}_t)$, see (2.99),¹¹ and it results into the minimum variance estimator (filter): $cov(\tilde{\mathbf{x}}_{\mathbf{P}^{-1}}) \leq cov(\hat{\mathbf{x}}_{\mathbf{W}})$ for all $\hat{\mathbf{x}}_t$ and \mathbf{W} ,
- if $\mathbf{W} = \mathbf{I}$, the least squares estimate of the state subject to state constraints is obtained and the constrained estimates is closer to the true state than the unconstrained one: $\|\mathbf{x}_t - \tilde{\mathbf{x}}_t\| \leq \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|$ for all t .

If the constrained forecasting step for the $\hat{\mathbf{x}}_{t+1|t}$ is based on the constrained estimate, then the constrained filter uses (2.52) adjusted (2.54) and (2.96) in the following way

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \tag{2.100}$$

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\tilde{\mathbf{x}}_{t|t} \tag{2.101}$$

$$\tilde{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t} - \mathbf{W}^{-1}\mathbf{D}^\top(\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^\top)^{-1}(\mathbf{D}\hat{\mathbf{x}}_{t|t} - \mathbf{d}). \tag{2.102}$$

If \mathbf{D} is a square matrix (the number of constraints is equal to the number of states) then the state vector is fully constrained. In this case, remembering the assumption that \mathbf{D} is full rank, equation (2.102) reduces to

$$\tilde{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t} - \mathbf{W}^{-1}\mathbf{D}(\mathbf{D}^{-\top}\mathbf{W}\mathbf{D}^{-1})(\mathbf{D}\hat{\mathbf{x}}_{t|t} - \mathbf{d}) = \mathbf{D}^{-1}\mathbf{d} \quad \Leftrightarrow \quad \mathbf{D}\tilde{\mathbf{x}}_{t|t} = \mathbf{d}. \tag{2.103}$$

¹¹At first sight, it seems counterintuitive, because the standard Kalman filter is by definition the minimum variance filter. However, the problem was changed by introducing state variable constraints. Therefore, the standard Kalman filter is no longer the minimum variance filter and better results are obtained by the constrained Kalman filter.

2.4.2 Kalman Gain Projection

Before the Kalman gain projection method is introduced, it is useful to recalculate matrix $\mathbf{P}_{t|t}$ in a different form compared to (2.53).

Kalman Gain Recalculation. The difference between the observed and predicted measurement is the predicted measurement error (measurement residual).

$$\mathbf{r}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t. \quad (2.104)$$

The associated covariance for the error, which is the expectation of the measurement error with itself, $\mathbb{E}(\mathbf{r}_t \mathbf{r}_t^\top)$, is (see equation (2.9))

$$\mathbf{S}_t = \mathbb{E}(\mathbf{r}_t \mathbf{r}_t^\top) = \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^\top + \mathbf{R}. \quad (2.105)$$

The updated state estimate in a form of (2.61), the prediction plus some perturbation, is given by a weighting factor times the forecast error. The weighting factor is the Kalman gain.

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \dot{\mathbf{K}}_t \mathbf{r}_t \quad (2.106)$$

$$\mathbf{r}_t = \mathbf{y}_t - \mathbf{A} \mathbf{u}_t - \mathbf{H} \hat{\mathbf{x}}_{t|t-1}. \quad (2.107)$$

The gain $\dot{\mathbf{K}}$ minimizes the mean square state estimate error, $\mathbb{E}|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}|^2$. This is the same as minimizing the trace of the updated error covariance matrix, because $\mathbf{r}_t \mathbf{r}_t^\top = \text{Tr}(\mathbf{r}_t \mathbf{r}_t^\top)$. The optimal gain, that achieves this, is (2.111).¹²

Recalculating a form for $\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}$ using (2.106), predicted measurement error (2.104) and its reformulation in a form of (2.8) give

$$\begin{aligned} \mathbf{x}_t - \hat{\mathbf{x}}_{t|t} &= \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1} + \dot{\mathbf{K}}_t \mathbf{r}_t \\ &= \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1} + \dot{\mathbf{K}}_t (\mathbf{y}_t - \hat{\mathbf{y}}_t) \\ &= \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1} + \dot{\mathbf{K}}_t [\mathbf{H}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{w}_t] \\ &= (\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \dot{\mathbf{K}}_t \mathbf{w}_t. \end{aligned} \quad (2.108)$$

Using the definition of the error covariance matrix (2.2) and $\mathbb{E}[\mathbf{w}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\top] = \mathbf{0}$ yield

$$\begin{aligned} \mathbf{P}_{t|t} &= \mathbb{E}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^\top \\ &= \mathbb{E} \left[(\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \dot{\mathbf{K}}_t \mathbf{w}_t \right] \left[(\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \dot{\mathbf{K}}_t \mathbf{w}_t \right]^\top \\ &= \mathbb{E} \left[(\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^\top (\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H})^\top \right] + \mathbb{E}(\dot{\mathbf{K}}_t \mathbf{w}_t \mathbf{w}_t^\top \dot{\mathbf{K}}_t^\top) \\ &= (\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H}) \mathbf{P}_{t|t-1} (\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H})^\top + \dot{\mathbf{K}}_t \mathbf{R} \dot{\mathbf{K}}_t^\top. \end{aligned} \quad (2.109)$$

Based on [72] and using the previous result, the Kalman filter can be derived by solving the following problem

$$\dot{\mathbf{K}}_t = \underset{\mathbf{K} \in \mathbb{S}\mathbb{S}}{\text{argmin}} \text{Tr}(\mathbf{P}_{t|t}) = \underset{\mathbf{K} \in \mathbb{S}\mathbb{S}}{\text{argmin}} \text{Tr} \left[(\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H}) \mathbf{P}_{t|t-1} (\mathbf{I} - \dot{\mathbf{K}}_t \mathbf{H})^\top + \dot{\mathbf{K}}_t \mathbf{R} \dot{\mathbf{K}}_t^\top \right]. \quad (2.110)$$

¹²The covariance matrices in the Kalman filter provide a measure for uncertainty in predictions and updated state estimate. This is an important feature for the various applications of filtering since it identifies how much to trust predictions and estimates.

The solution to this problem gives the optimal Kalman gain in a form of equation (2.60)

$$\dot{\mathbf{K}}_t = \mathbf{P}_{t|t-1} \mathbf{H}^\top (\mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^\top + \mathbf{R})^{-1} = \mathbf{P}_{t|t-1} \mathbf{H}^\top \mathbf{S}_t^{-1}, \quad (2.111)$$

where \mathbf{S}_t is expressed as (2.105).

Imposing Inequality Constraints. The gain projection method was applied to the inequality constraints in [76]. The method solves the minimization problem of the gain projection (2.110) – a Kalman gain is found such that the updated state estimate will be forced to lie in the constrained space. However, the method will no longer be able to be found an analytic solution compared to a case of the equality constraints, using the method of the Lagrange multipliers, see [31].

This problem can be solved by restricting the optimal Kalman gain – we seek the optimal $\tilde{\mathbf{K}}_t$ that satisfies the constrained optimization problem

$$\tilde{\mathbf{K}}_t = \underset{\mathbf{K} \in \mathcal{SS}}{\operatorname{argmin}} \operatorname{Tr} \left[(\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}_{t|t-1} (\mathbf{I} - \mathbf{K} \mathbf{H})^\top + \mathbf{K} \mathbf{R} \mathbf{K}^\top \right]. \quad (2.112)$$

such that

$$\mathbf{D} \hat{\mathbf{x}}_{t|t} = \mathbf{D} (\hat{\mathbf{x}}_{t|t-1} + \dot{\mathbf{K}}_t \mathbf{r}_t) \leq \mathbf{d}. \quad (2.113)$$

This can be solved using any inequality constrained optimization method.

Constraints for the State Forecasting. If the estimate of $\hat{\mathbf{x}}_{t+1|t}$ satisfies the constraints, but the unconstrained update estimate $\hat{\mathbf{x}}_{t|t}$ does not satisfy them, then $\hat{\mathbf{x}}_{t+1|t}$ can be projected in the direction of $\hat{\mathbf{x}}_{t|t}$ until it reaches the constraint boundary at each forecasting step of the state vector. This finally gives a modification of the unconstrained Kalman gain $\dot{\mathbf{K}}_t$ in the form of

$$\dot{\mathbf{K}}_t^{\text{modified}} = \beta_t \dot{\mathbf{K}}_t, \quad (2.114)$$

where $\beta_t \in (0; 1)$ for all t .

2.4.3 Probability Density Function Truncation

In the probability density function truncation approach, the PDF of the state estimate, which is computed by the Kalman filter, is taken, assuming that it is Gaussian, and truncated at the constraint edges. The constrained state estimate is equal to the mean of the truncated PDF. This approach is designed for inequality constraints on the state. It can also be applied to equality constraints with some modifications.

Figure 2.3 shows an example of a one-dimensional state estimate before and after truncation. Before truncation the state estimate is outside of the state constraints. After truncation, the state estimate is set equal to the mean of the truncated PDF. The Gaussian PDF is truncated at the constraint boundaries, and the constrained estimate is equal to the mean of the truncated PDF. An initial consideration of the left picture might indicate that the constrained estimate should lie on the constraint boundary. In fact, this is exactly the philosophy of the projection approach to constrained filtering, see Subsection 2.4.1. However, the PDF truncation approach considers both the constraints and the unconstrained Kalman filter's Gaussian distribution of the estimate. The resulting constrained estimate lies at a place within the constraint boundaries that is determined by both the information from the unconstrained filter and the constraints.

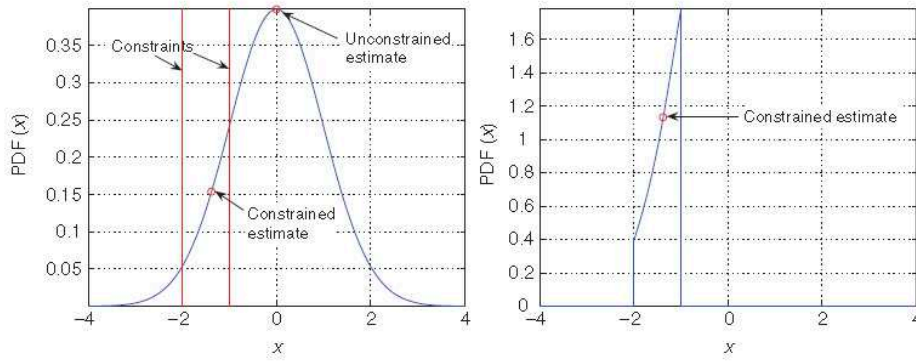


Figure 2.3: Unconstrained and Constrained PDF - The Unconstrained Estimate Does Not Satisfy the Constraint (Source: [74])

Figure 2.4 shows another example of the PDF truncation. In this case the unconstrained state estimate is inside the state constraints. However, truncation changes the PDF and so the constrained state estimate changes to the mean of the truncated PDF. It could be argued that the estimate should not be changed if it satisfies the constraints. In fact, the PDF truncation filter could be implemented either way. Whether to modify estimates that already satisfy the constraints (as shown in Figure 2.4), or leave those estimates unchanged, is an implementation decision that depends on the application and a judgment.

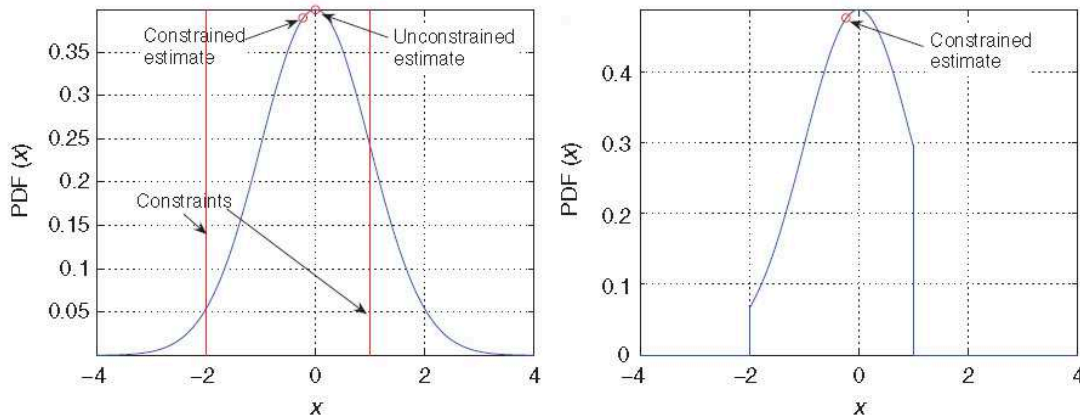


Figure 2.4: Unconstrained and Constrained PDF - The Unconstrained Estimate Satisfies the Constraints (Source: [74])

This method is complicated when the state dimension is more than one. In this case the state estimate is normalized using an assumption that the components of the state estimation are statistically independent of each other. Then the normalized constraints are applied together at a time. After all the constraints are applied, the normalization process is reversed to obtain the constrained state estimate.

On top of that, this method imposes a bias on the state estimate. The unconstrained Kalman filter has the property that the state estimate is the mean of the true state conditioned on the measurements. However, the truncated state

estimate is biased. This is a drawback to this method of constraint enforcement, especially since other methods of constraint enforcement preserve unbiasedness. However, if other features of the estimate are more important to the user than unbiasedness (e.g. root mean square estimation error) then the truncation approach to constraint enforcement may remain attractive.

Details about the PDF truncation algorithm for equality and inequality constraints are given in [72] and especially in [74]. Although this algorithm was introduced only for linear state constraints, it is not conceptually difficult to extend to non-linear constraints. If the state constraints are non-linear, they can be linearized. The PDF truncation method has been extended to non-linear unscented Kalman filters in [81].

2.4.4 State Prediction Constraints

Forcing the constraints on the state equation should provide a better prediction. Ideally, the transition matrix \mathbf{F} is time-varying, see Subsection 2.3.2, and will take an updated state estimate satisfying the constraints at time $t - 1$ and make a prediction that will satisfy the constraints at time t . Of course this may not be the case. In fact, the constraints may depend on the updated state estimate, which would be the case for non-linear constraints. Constraining the state prediction increases computational cost per iteration.

For a simplification, the prediction of the transition time-varying matrix of the state equation is labeled $\mathbf{F}_{t|t+1}$.

There are three methods which are commonly used, see [31]. On top of that, different approach for state prediction constraints was introduced within the Kalman gain projection method.

Projection of Matrix $\mathbf{F}_{t|t+1}$ onto the Constrained Space. This method is feasible for the equality constraints only, as there is no trivial way to project $\mathbf{F}_{t|t+1}$ to an inequality constrained space. The same projector as in (2.99) is used

$$\tilde{\mathbf{F}}_{t|t-1} = (\mathbf{I} - \mathbf{CD})\mathbf{F}_{t|t-1}. \quad (2.115)$$

This new transition matrix will make a prediction that will keep the estimate in the equality constrained space.

Minimization Problem. If the previous assumptions are weakened, i.e., we are not constraining the updated state estimate, we could solve the minimization problem analogous to (2.95). Then it is possible to incorporate inequality constraints. In this situation the covariance matrix is constrained, similarly to the estimate projection method, see Subsection 2.4.1.

Additional Constraints. The idea of this method is to add to the constrained problem the additional constraints, which ensure that the chosen estimate will produce a prediction at the next iteration that is also constrained. This allows to have time-varying matrix \mathbf{D}_t and vector \mathbf{d}_t . To solve the problem, an iterative method is employed for the constraints

$$\mathbf{D}_t\mathbf{F}_{t|t-1}\mathbf{x}_t = \mathbf{d}_t \quad \mathbf{D}_t\mathbf{F}_{t|t-1}\mathbf{x}_t \leq \mathbf{d}_t. \quad (2.116)$$

2.4.5 Soft Constraints

Soft constraints, as opposed to hard constraints (which were discussed so far), are constraints that are only required to be approximately satisfied rather than exactly satisfied. They can be implemented in the Kalman filter in various ways

- the perfect measurement approach, where state equality constraints are treated as perfect measurements with zero measurement error, can be extended to soft constraints by adding small non-zero measurement noise to the perfect measurements, see e.g. [55] or [86],
- soft constraints can be implemented by adding a regularization term to the standard Kalman filter, see [74],
- soft constraints can be enforced by projecting the unconstrained estimates in the direction of the constraints rather than exactly onto the constraint surface, see [56].

The incorporation of state variable constraints increases the computational effort of the filter, but significantly improves its estimation accuracy. Additionally, [74] shows that the algorithms with soft constraints provide improved performance over unconstrained Kalman filtering.

Note that, if \mathbf{d} is uncertain, then the constraints can be replaced by a noisy inequality constraint, see e.g. [44]

$$\mathbf{D}\mathbf{x}_t + \rho \leq \mathbf{d}. \quad (2.117)$$

Noise ρ is generally assumed to be a zero-mean Gaussian random vector with appropriate covariance which reflects the level of uncertainty in the constraint. By considering ρ as a measurement noise, the unconstrained Kalman filter estimate can be modified to which the constraint is not yet applied. Since the constraint is used as a measurement, this method is also termed the pseudo-measurement method.¹³

Adding a Regularization Term. This approach uses a Kalman filter-based state estimate for state variables which we know a priori vary slowly with time. To simplify the situation, let's assume that matrix \mathbf{F} in state equation (1.13) is an identity matrix.

We can use the results of the estimate projection, especially (2.92), to formulate the problem for all t

$$\min_{\tilde{\mathbf{x}}_t \in \mathbf{S}} (\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_{t|t})^\top \mathbf{W} (\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_{t|t}) \quad \text{such that } \tilde{\mathbf{x}}_t \text{ varies slowly}, \quad (2.118)$$

where, similarly as before, \mathbf{W} is a constant symmetric positive definite weighting matrix. This is a type of regularization: some additional structure is incorporated into the Kalman filter estimate, see e.g. [71]. The problem can be reformulated as

$$\min_{\tilde{\mathbf{x}}_t \in \mathbf{S}} \left[(\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_{t|t})^\top \mathbf{W} (\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_{t|t}) + (\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_{t-1})^\top \mathbf{V}_t (\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_{t-1}) \right], \quad (2.119)$$

¹³In case of the inequality constraints, they can be replaced by the equality ones with a specific properties of the measurement noise ρ .

where \mathbf{V}_t is a (possibly time-varying) symmetric positive definite weighting matrix that balances the desire for a close approximation to $\tilde{\mathbf{x}}_t$ and smooth estimate $\hat{\mathbf{x}}_{t|t}$. The solution¹⁴ to the previous problem is

$$\tilde{\mathbf{x}}_0 = \mathbb{E}\mathbf{x}_0 \tag{2.120}$$

$$\tilde{\mathbf{x}}_t = (\mathbf{W} + \mathbf{V}_t)^{-1}(\mathbf{W}\hat{\mathbf{x}}_{t|t} + \mathbf{V}_t\tilde{\mathbf{x}}_{t-1}). \tag{2.121}$$

The soft constrained estimated state has the following properties, see [74]

- the solution $\tilde{\mathbf{x}}_t$ is an unbiased state estimator for any symmetric positive definite weighting matrices \mathbf{W} and \mathbf{V}_t : $\mathbb{E}\tilde{\mathbf{x}}_t = \mathbb{E}\mathbf{x}_t$ for all t ,
- the constrained state estimate approaches the unconstrained estimate in the limit as time goes to infinity, assuming that $\mathbf{v}_t = 0$: $\lim_{t \rightarrow \infty} \tilde{\mathbf{x}}_t = \lim_{t \rightarrow \infty} \hat{\mathbf{x}}_{t|t}$,
- if $\mathbf{V}_t = (t-1)\mathbf{W}$ in (2.121) then $\tilde{\mathbf{x}}_t$ is the arithmetic average of $\hat{\mathbf{x}}_{i|i}$. The (arithmetic) average of $\tilde{\mathbf{x}}_t$ is defined simply as $X_t = \frac{1}{t} \sum_{i=1}^t \hat{\mathbf{x}}_{i|i}$, which implies that $X_{t+1} = \frac{1}{t+1}(\hat{\mathbf{x}}_{t+1|t+1} + tX_t)$. If $\mathbf{V}_t = (t-1)\mathbf{W}$ then using (2.121) implies $\tilde{\mathbf{x}}_t = ((t+1)\mathbf{W})^{-1}(\mathbf{W}\hat{\mathbf{x}}_{t+1|t+1} + t\mathbf{W}\tilde{\mathbf{x}}_t) = \frac{1}{t+1}(\hat{\mathbf{x}}_{t+1|t+1} + t\tilde{\mathbf{x}}_t)$, which is the average as defined earlier.

¹⁴Because \mathbf{W} and \mathbf{V}_t are both positive definite, $(\mathbf{W} + \mathbf{V}_t)^{-1}$ exists.

3. The Non-Linear Kalman Filter and/or Non-Linear Constraints

So far, we have dealt with linear models and constraints in the Kalman filter. Traditionally, problems using the Kalman filter are complex and non-linear. Many advances have been made in the direction of dealing with non-linearities (e.g., the extended or unscented Kalman filter). These problems also tend to have inherent state-space inequality constraints.

The linear models (and constraints) are often employed for an approximated description of the non-linear controlled system (or constraints) due to the simplicity of identification algorithms. However, these models have only limited validity. The non-linear models describe the system generally much better than the linear ones but their identification is a non-trivial task.

3.1 Non-Linear Kalman Filters

As described earlier in Chapter 2, the Kalman filter addresses the general problem of trying to estimate the state of a discrete-time controlled process that is governed by a linear stochastic difference equation. If the process or/and the measurement relationship to the process is non-linear, it is necessary either to adjust the Kalman filter or find another method.

In this section, a non-linear state-space model is considered in a form of (1.22) and (1.23).

3.1.1 The Extended Kalman Filter

The method proposed by the extended Kalman filter (EKF) is to linearize the non-linearities around the current state prediction (or estimate). That is, we choose \mathbf{F}_t as the Jacobian of $\mathbf{f}(\mathbf{x}_t)$ evaluated at $\hat{\mathbf{x}}_{t|t}$, and \mathbf{H}_t as the Jacobian of $\mathbf{h}(\mathbf{x}_t)$ evaluated at $\hat{\mathbf{x}}_{t|t-1}$ and proceed as the linear Kalman filter.¹

In this situation a fundamental flaw of the EKF is that the distributions (or densities in the continuous case) of the various random variables are no longer normal after undergoing their respective non-linear transformations. The EKF is simply an ad hoc state estimator that only approximates by the first order linearization. Some interesting work has been done in developing a variation to the EKF, using methods that preserve the normal distributions throughout the non-linear transformations, see [39]. Additionally, in practice, the use of the EKF has two drawbacks, see [39]

- linearization can produce highly unstable filters if the assumptions of local linearity are violated,
- the derivation of the Jacobian matrices are non-trivial in most applications and often lead to significant implementation difficulties.

¹Expanding some vector function $\mathbf{f}(\mathbf{x})$ in Taylor series about \mathbf{x}^0 yields $\mathbf{f}(\mathbf{x}) \equiv \mathbf{f}(\mathbf{x}^0) + \mathbf{J}_f(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0) + \dots$, where \mathbf{J}_f is the Jacobian of $\mathbf{f}(\mathbf{x})$ and the higher order terms are considered negligible. Hence, the extended Kalman filter is also called the first-order filter, see [83].

To estimate a process with non-linear difference and measurement relationships, we begin by writing new governing equations that linearize an estimate about (1.22) and (1.23). One iteration of the EKF is composed by the following consecutive steps, see [67]

- consider the last filtered state estimate $\hat{\mathbf{x}}_{t|t}$,
- linearize the system dynamics $\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t) + \mathbf{v}_{t+1}$ around $\hat{\mathbf{x}}_{t|t}$,
- apply the prediction step of the Kalman filter to the linearized system dynamics just obtained, yielding $\hat{\mathbf{x}}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$,
- linearize the observation dynamics $\mathbf{y}_t = \mathbf{a}(\mathbf{u}_t) + \mathbf{h}(\mathbf{x}_t) + \mathbf{w}_t$ around $\hat{\mathbf{x}}_{t+1|t}$,
- apply the filtering or update cycle of the Kalman filter to the linearized observation dynamics, yielding $\hat{\mathbf{x}}_{t+1|t+1}$ and $\mathbf{P}_{t+1|t+1}$.

It is important to state that the EKF is not an optimal filter, but rather it is implemented based on a set of approximations. Thus, matrices $\mathbf{P}_{t|t}$ and $\mathbf{P}_{t+1|t}$ do not represent the true covariance of the state estimates. Moreover, contrary to the linear Kalman filter, the EKF may diverge, if the consecutive linearizations are not a good approximation of the linear model in all associated uncertainty domain.

For a detailed description and a derivation of the equations for the EKF see [67] or [83].

3.1.2 The Unscented Kalman Filter

The unscented Kalman filter (UKF) is a filter for non-linear systems which is based on an idea that the unscented transformation is easier to approximate a Gaussian distribution than to approximate an arbitrary non-linear function or transformation. A set of points (sigma points) are chosen so that their sample mean and covariance remain unchanged to the original PDF. For a further description of the algorithm and equations see [81].

The previous implies also an assumption that it is not difficult to find a set of vectors (sigma points) in state-space whose sample PDF approximates a given PDF, see [39]. The UKF uses between $(r + 1)$ and $(2r + 1)$ sigma points, where r is the dimension of the state vector. The sigma points are transformed and combined in a special way in order to obtain an estimate of the state and an estimate of the covariance of the state estimation error. For a detailed discussion about this transformation and its application to the Kalman filter see [39].

It means that instead of analytically or numerically linearizing non-linear state and measurement equations, the UKF employs the unscented transformation (UT), which approximates the posterior mean and covariance matrix of a random vector \mathbf{y}_T obtained from the non-linear transformation of the measurement equation (1.23). The UT yields the true mean and the true covariance of \mathbf{y}_T if $\mathbf{h}(\mathbf{x}_t) = \mathbf{h}_1(\mathbf{x}_t) + \mathbf{h}_2(\mathbf{x}_t)$, where $\mathbf{h}_1(\mathbf{x}_t)$ is linear and $\mathbf{h}_2(\mathbf{x}_t)$ is quadratic. Otherwise, we speak about a pseudo mean and a pseudo covariance matrix. For more details see [38].

3.1.3 Modifications of the Unscented Kalman Filter

To complete the survey in this area, here is only a brief introduction of basic methods among a bunch of modification of the UKF, based on [73]. For a detailed discussion see the quoted references.

One possibility is to base the forecasted state estimate on the unconstrained UKF with updated state estimate from the previous time step, [81] and [82]. In this case the standard unconstrained UKF runs independently of the constrained UKF. At each time the state estimate of the unconstrained UKF is combined with the constraints, which are treated as perfect measurements, to obtain a constrained a posteriori UKF estimate. This filter is referred to as the projected UKF (PUKF) and is analogous to (2.100) – (2.102) for linear systems and constraints in case of the estimate projection, see Subsection 2.4.1.

Another way is to base the forecasted state estimate on the constrained UKF with updated state estimate from the previous time step [81]. At each time the state estimate of the unconstrained UKF is combined with the constraints, which are treated as perfect measurements, to obtain a constrained a posteriori UKF estimate. This constrained updated estimate is then used as the initial condition for the next time update. This filter is referred to as the equality constrained UKF (EUKF) and is also identical to the measurement-augmentation UKF in [81]. The ECUKF is analogous to results for the estimate projection with inequality constraints for linear systems and linear constraints. A similar filter is explored in [36], where it is argued that the covariance of the constrained estimate is expected to be larger than that of the unconstrained estimate since the unconstrained estimate approximates the minimum variance estimate.

The two-step UKF (2UKF) in [36] projects each updated sigma point onto the constraint surface to obtain constrained sigma points. The state estimate is obtained by taking the weighted mean of the sigma points in the usual way and the resulting estimate is then projected onto the constraint surface. Note that the mean of constrained sigma points does not necessarily satisfy a non-linear constraint. 2UKF is unique in that the estimation error covariance increases after the constraints are applied. The argument for this increase is that the unconstrained estimate is the minimum variance estimate, so changing the estimate by applying constraints should increase the covariance. Furthermore, if the covariance decreases with the application of constraints (e.g. using the algorithms in [75] and [82]) then the covariance might become singular, which might lead to numerical problems with the matrix square root algorithm of the unscented transformation.

Unscented recursive non-linear dynamic data reconciliation (URNDDR), see [84], is similar to 2UKF. URNDDR projects the updated sigma points onto the constraint surface, and modifies their weights based on their distances from the a posteriori state estimate. The modified updated sigma points are passed through the dynamic system in the usual way to obtain the a priori sigma points at the next time step. The next set of a posteriori sigma points is obtained using a non-linear constrained MHE with a horizon size of one. This approach requires the solution of a non-linear constrained optimization problem for each sigma point. The updated state estimate and covariance are obtained by combining the sigma points in the normal way. The constraints are thus used in two different ways for

the a posteriori estimates and covariances.²

The constrained UKF (CUKF) is identical to the standard UKF, see [82]. Sigma points are not projected onto the constraint surface, and constraint information is not used to modify covariances. The constrained interval UKF (CIUKF) combines the sigma point constraints of URNDDR with the measurement update of the CUKF. That is, the CIUKF is the same as URNDDR except instead of using MHE to constrain the updated sigma points, the unconstrained sigma points are combined to form an unconstrained estimate, and then MHE is used to constrain the estimate.

The interval UKF (IUKF) combines the post-measurement projection step of URNDDR with the measurement update of the standard unconstrained UKF, see [82]. That is, the IUKF is the same as URNDDR except that it skips the MHE-based constraint of the a posteriori sigma points.

The truncated UKF (TUKF) combines the PDF truncation approach, described in Subsection 2.4.3. After each measurement update of the UKF, the PDF truncation approach is used to generate a constrained state estimate and covariance. The constrained estimate is used as the initial condition for the following time update. The truncated interval UKF (TIUKF) adds the PDF truncation step to the a posteriori update of the IUKF. As with the TUKF, the constrained estimate is used as the initial condition for the following time update, see [82].

3.2 Constraints

Consider the non-linear system of (1.22) and (1.23) with state linear constraints. The constraints can be applied in the same way as in the linear Kalman filter, see Section 2.4. In case that the state constraints are non-linear, it is used instead of (2.84) in the form of $\mathbf{D}\mathbf{x}_t \leq \mathbf{d}$ the following specification

$$\mathbf{g}(\mathbf{x}_t) \leq \mathbf{h}. \quad (3.1)$$

Then a Taylor series expansion of the constraint equation around $\hat{\mathbf{x}}_{t+1|t}$ is used

$$\begin{aligned} \mathbf{g}(\mathbf{x}_t) &\approx \mathbf{g}(\hat{\mathbf{x}}_{t+1|t}) + \mathbf{g}'(\hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t+1|t}) \\ &\quad + \frac{1}{2} \sum_{i=1}^s \mathbf{e}_i (\mathbf{x}_t - \hat{\mathbf{x}}_{t+1|t})^\top \mathbf{g}''(\hat{\mathbf{x}}_{t+1|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t+1|t}), \end{aligned} \quad (3.2)$$

where s is the dimension of $\mathbf{g}(\mathbf{x}_t)$, \mathbf{e}_i is the i -th canonical basis vector in \mathbb{R}^s , and the entry in the p -th row and q -th column of the $(n \times n)$ matrix of the second partial derivatives $\mathbf{g}''_i(\mathbf{x})$ is given by

$$[\mathbf{g}''_i(\mathbf{x})]_{pq} = \frac{\partial^2 \mathbf{g}_i(\mathbf{x})}{\partial \mathbf{x}_p \partial \mathbf{x}_q}, \quad (3.3)$$

for $i = 1, \dots, s$ and $p, q = 1, \dots, r$. Similarly for the first partial derivatives, where $[\mathbf{g}'_i(\mathbf{x})]_p = \frac{\partial \mathbf{g}_i(\mathbf{x})}{\partial \mathbf{x}_p}$.

²URNDDR is called the sigma point interval UKF in [82].

In the following subsections there is a brief tour through the most used methods for non-linear state constraints handling. This list is not complete and states only commonly used techniques.

3.2.1 Second-Order Expansion

This method allows for the use of second-order non-linear state constraints. It can provide better approximation to higher order of non-linearities. Considering only second-order constraints is a tradeoff between reducing approximation errors to higher-order non-linearities and keeping the problem computationally tractable.

An idea behind the second-order expansion is similar to the extended Kalman filter, which relies on linearization of the system and measurement equations. It can be improved by retaining second-order terms to obtain the second-order extended Kalman filter, see e.g. [72]. This means that the second-order term in the Taylor expansion of the constraint in (3.2) is kept and the EKF is run.

3.2.2 Moving Horizon Estimation

Moving horizon estimation (MHE) is another optimization approach that uses a series of measurements observed over time, containing noise (random variations) and other inaccuracies, and produces estimates of unknown variables or parameters. Unlike deterministic approaches like the Kalman filter, MHE requires an iterative approach that relies on linear programming or non-linear programming solvers to find a solution.

In order to obtain a tractable solution, MHE relies on the assumption of Gaussian prior and noise to obtain a least-squares estimation problem. However, a closed-form solution is not available anymore. Instead, MHE needs to solve a constrained optimization problem over each moving window. In addition, MHE lacks an accurate and fast algorithm for propagating the posterior, thus fails to enjoy a full recursive formulation.

MHE reduces to the Kalman filter under certain simplifying conditions. A critical evaluation of the extended Kalman filter and MHE found improved performance of MHE with the only cost of improvement being the increased computational expense. MHE is attractive in the generality of its formulation. But because of the computational expense, MHE is usually applied to systems where there are greater computational resources.

For further references see [73], or in more details [65] and [66].

3.2.3 Interior Point Method

A relatively new approach to inequality-constrained state estimation is called interior point likelihood maximization (IPLM), introduced in [7]. This approach is based on interior point methods, which are fundamentally different from active set methods for constraint enforcement.

One difficulty with active set methods is that computational effort grows exponentially with the number of constraints. Interior point approaches solve inequality-constrained problems by iterating using a Newton's method that is applied to a certain subproblem. The IPLM approach also relies on linearization.

4. State Uniform Model Filtration

In the previous chapters, there was an assumption that the noises of state evolution and observations have a normal distribution and the problem is then solved by means of the Kalman filter. This can cause difficulties and there are several ways how to deal with this drawback. One of a possible solution is a discrete-time state-space model with uniformly distributed state and measurement noises which is introduced in this chapter. For the state estimation, the Bayesian approach is applied.

This chapter is based on [62] and [64] if it is not emphasized otherwise.¹ An extension to a non-linear system is presented in [61].

4.1 State-Space Model with Bounded Noise

A discrete-time linear state-space model, defined by (1.13) and (1.15), is used for this chapter. The vectors of the state and measurement (output) noises \mathbf{v}_t and \mathbf{w}_t are zero mean with constant variances, mutually conditionally independent and identically distributed. Additionally, they are assumed to have uniform distribution on the multivariate boxes with the $\mathbf{0}$ center and half-widths of the support interval equal to \mathbf{p} and \mathbf{q} , respectively

$$\mathbf{v}_t \sim \mathcal{U}(\mathbf{0}, \mathbf{p}) \qquad \mathbf{w}_t \sim \mathcal{U}(\mathbf{0}, \mathbf{q}), \quad (4.1)$$

where $\mathbf{0}$ is a vector of zeros of an appropriate size. These assumptions and using the state-space representation of the model imply

$$(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t, \mathbf{p}, \mathbf{q}) \sim \mathcal{U}(\mathbb{E}\mathbf{x}_{t+1}, \mathbf{p}) \qquad (\mathbf{y}_t | \mathbf{x}_t, \mathbf{u}_t, \mathbf{p}, \mathbf{q}) \sim \mathcal{U}(\mathbb{E}\mathbf{y}_t, \mathbf{q}). \quad (4.2)$$

Further, \mathbf{x}_0 , \mathbf{p} , and \mathbf{q} are supposed to be a priori mutually independent and restricted

$$\underline{\mathbf{x}}_0 \leq \mathbf{x}_0 \leq \bar{\mathbf{x}}_0 \qquad \underline{\mathbf{x}} \leq \mathbf{x}_t \leq \bar{\mathbf{x}} \quad (4.3)$$

$$\mathbf{0} \leq \mathbf{p} \leq \bar{\mathbf{p}} \qquad \mathbf{0} \leq \mathbf{q} \leq \bar{\mathbf{q}} \quad (4.4)$$

These restrictions are defined by the user so that they reflect the reality and specify user's prior information.

State-space model equations (1.13) and (1.15) together with the uniform distribution assumptions (4.1) and restrictions (4.3) and (4.4) define the state uniform (SU) model.

The introduced SU model provides the following advantages

- it respects natural bounds on stochastic disturbances,
- it allows estimation of the innovation range, and

¹Similar idea about an approximate estimation of autoregressive model with exogenous variables (ARX model) with a uniform noise is introduced in [42].

- it allows to respect prior bounds on model parameters and states.

To simplify a notation, the column vector \mathbf{X} consists of

$$\mathbf{X} = [(\mathbf{xx}_T)^\top \quad \mathbf{p}^\top \quad \mathbf{q}^\top]^\top. \quad (4.5)$$

The joint PDF (1.38) of data \mathbf{uy}_T , the state trajectory \mathbf{xx}_T and unknown parameters $\theta = [\mathbf{p}^\top, \mathbf{q}^\top]^\top$ of the SU model takes the form

$$p(\mathbf{uy}_T, \mathbf{xx}_T, \theta) \propto \left(\prod_{i=1}^r p_i \prod_{j=1}^n q_j \right)^{-T} \mathbb{I}_{\mathcal{S}}, \quad (4.6)$$

where r and n are the lengths of the state and measurement vector, respectively, and $\mathbb{I}_{\mathcal{S}}$ is the indicator of the support \mathcal{S} of this PDF,

$$\mathcal{S} = \mathcal{S}_0 \cap \mathcal{S}_1 \cap \mathcal{S}_2, \quad (4.7)$$

where restrictions on the state values, based on (4.3), are in the following form

$$\mathcal{S}_2 = \{\underline{\mathbf{x}}_t \leq \mathbf{x}_t \leq \bar{\mathbf{x}}_t\}_{t=1}^T. \quad (4.8)$$

Further \mathcal{S}_1 is a set of \mathbf{X} that fulfills state-space model equations (1.13) and (1.15) at each time point t , i.e.,

$$\mathcal{S}_1 = \{-\mathbf{p} \leq \mathbf{x}_{t+1} - \mathbf{F}\mathbf{x}_t \leq \mathbf{p}; \quad -\mathbf{q} \leq \mathbf{y}_t - \mathbf{A}\mathbf{u}_t - \mathbf{H}\mathbf{x}_t \leq \mathbf{q}\}_{t=1}^T \quad (4.9)$$

and \mathcal{S}_0 is a set of \mathbf{X} that meets the first inequality in (4.3) and innovations bounds (4.4)

$$\mathcal{S}_0 = \{\underline{\mathbf{x}}_0 \leq \mathbf{x}_0 \leq \bar{\mathbf{x}}_0; \quad \mathbf{0} \leq \mathbf{p} \leq \bar{\mathbf{p}}; \quad \mathbf{0} \leq \mathbf{q} \leq \bar{\mathbf{q}}\}. \quad (4.10)$$

They are assumed a priori mutually independent, hence

$$p(\mathbf{x}_t, \mathbf{p}, \mathbf{q}) = p(\mathbf{x}_t)p(\mathbf{p})p(\mathbf{q}) \quad (4.11)$$

yielding into (4.6).

This approach is possible to further generalize by introducing partially unknown system matrices in the state-space model representation, see [64].

4.2 Linear State and Noise Boundary Estimation

The required posterior PDF can be calculated according to Bayes' rule. However, the number of vertices of the support \mathcal{S} is proportional to the number of data, which is a large number for realistic situations. Consequently, evaluation of moments of this PDF is computationally demanding. This is why the maximum a posteriori probability (MAP) estimate is evaluated, see e.g. [9].

The method of the MAP estimation provides a point estimate of an unobserved \mathbf{X} on the basis of observed data \mathbf{uy}_T and prior information about \mathbf{X} ,

$$\hat{\mathbf{X}} = \operatorname{argmax}_{\mathbf{X} \in \mathcal{X}} \frac{p(\mathbf{uy}_T | \mathbf{X})p(\mathbf{X})}{\int_{\mathcal{X}} p(\mathbf{uy}_T | \mathbf{X})p(\mathbf{X})d\mathbf{X}} = \operatorname{argmax}_{\mathbf{X} \in \mathcal{X}} p(\mathbf{uy}_T | \mathbf{X})p(\mathbf{X}). \quad (4.12)$$

The denominator of the posterior distribution does not depend on \mathbf{X} and therefore plays no role in the optimization.

Using the system described by the SU model (4.6), vector \mathbf{X} (that is states \mathbf{xx}_T and noise bounds \mathbf{p} and \mathbf{q}) is estimated

$$\hat{\mathbf{X}} = \operatorname{argmax}_{\mathbf{X} \in \mathcal{S}} \left(\prod_{i=1}^r (p_i)^{-T} + \prod_{j=1}^n (q_j)^{-T} \right). \quad (4.13)$$

The support \mathcal{S} defines the maximization domain. For further calculation, the MAP estimate $\hat{\mathbf{X}}$ of \mathbf{X} is used with negative linearized logarithm of a posteriori PDF and divided by T

$$\hat{\mathbf{X}} = \operatorname{argmin}_{\mathbf{X} \in \mathcal{S}} \left(\sum_{i=1}^r \log(p_i) + \sum_{j=1}^n \log(q_j) \right). \quad (4.14)$$

Using the Taylor approximation for the logarithm of p_i and q_i , we obtain

$$\hat{\mathbf{X}} = \operatorname{argmin}_{\mathbf{X} \in \mathcal{S}} \left(\sum_{i=1}^r p_i + \sum_{j=1}^n q_j \right). \quad (4.15)$$

Thus, the state and parameter estimations become a linear programming task, for all t , see [27]

$$\min_{\mathbf{X} \in \mathcal{S}} \mathbf{a}^\top \mathbf{X} = \sum_{i=1}^r p_i + \sum_{j=1}^n q_j \quad (4.16)$$

such that

$$\mathbf{C}\mathbf{X} \leq \mathbf{b}, \quad (4.17)$$

$$\underline{\mathbf{X}} \leq \mathbf{X} \leq \bar{\mathbf{X}}, \quad (4.18)$$

where

- vector \mathbf{a} is

$$\mathbf{a}^\top \equiv [\mathbf{0}^\top, \mathbf{1}^\top], \quad (4.19)$$

and $\mathbf{0}$ and $\mathbf{1}$ are the vectors of zeros and ones of the following lengths ($n(T+1)$) and $(r+n)$ respectively,

- \mathbf{C} is a known matrix and \mathbf{b} is a known vector. They result from the inequalities describing the set \mathcal{S}_1 by equation (4.9). They are reorganized so that terms containing entries from \mathbf{X} are on the left-side and have the form of (4.17), where

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \quad (4.20)$$

and

$$\mathbf{C}_{11} = \mathcal{L}_r (\mathbf{I}_{(T)} \otimes \mathbf{K} \otimes \mathbf{I}_{(r)}) - \mathcal{R}_r (\mathbf{I}_{(T)} \otimes \mathbf{K} \otimes \mathbf{F}) \quad (4.21)$$

$$\mathbf{C}_{12} = -\mathbf{1}_{(2T)} \otimes \mathcal{R}_n (\mathbf{I}_{(r)}) \quad (4.22)$$

$$\mathbf{C}_{21} = \mathcal{L}_r (\mathbf{I}_{(T)} \otimes \mathbf{K} \otimes \mathbf{H}) \quad (4.23)$$

$$\mathbf{C}_{22} = -\mathbf{1}_{(2T)} \otimes \mathcal{L}_r (\mathbf{I}_{(n)}) \quad (4.24)$$

$$\mathbf{b}_1 = \mathbf{0}_{(2T,1)} \quad (4.25)$$

$$\mathbf{b}_2 = [\mathbf{I}_{(T)} \otimes \mathbf{K} \otimes \mathbf{I}_{(n)}] \mathbf{y}\mathbf{y}_T - [\mathbf{I}_{(T)} \otimes \mathbf{K} \otimes \mathbf{A}] \mathbf{u}\mathbf{u}_T, \quad (4.26)$$

where symbol \otimes denotes the Kronecker product, $\mathcal{L}_{col}(\cdot)$ and $\mathcal{R}_{col}(\cdot)$ are operators adding zero columns of col length to the specified matrix from the left or right,

- $\underline{\mathbf{X}}$ and $\overline{\mathbf{X}}$ are known vectors - they stem from the set \mathcal{S}_0 and \mathcal{S}_2 , defined by equations (4.8) and (4.10), and have the following forms

$$\underline{\mathbf{X}} = \begin{bmatrix} \underline{\mathbf{x}}_0 \\ \mathbf{1}_{(2rT)} \otimes \underline{\mathbf{x}} \\ \mathbf{0}_{(r,1)} \\ \mathbf{0}_{(n,1)} \end{bmatrix} \quad \overline{\mathbf{X}} = \begin{bmatrix} \overline{\mathbf{x}}_0 \\ \mathbf{1}_{(2rT)} \otimes \overline{\mathbf{x}} \\ \overline{\mathbf{p}} \\ \overline{\mathbf{q}} \end{bmatrix}. \quad (4.27)$$

In general, the parameter estimation can be done by two possibilities

- off-line state and parameter estimation, where one shot estimation of states \mathbf{x}_t for $t = 1, \dots, T$ and the innovation boundaries \mathbf{p} and \mathbf{q} is run, or
- real-time (on-line) estimation which provides the state and/or parameter estimates in each time step, once new data is available.

For a detailed description see [64].

Estimation of the State. Matrix \mathbf{C} and vector \mathbf{b} are given by the inequalities describing the set \mathcal{S}_1 rearranged into the following form

$$\mathbf{x}_{t+1} - \mathbf{F}\mathbf{x}_t - \mathbf{p} \leq \mathbf{0} \quad (4.28)$$

$$-\mathbf{x}_{t+1} + \mathbf{F}\mathbf{x}_t - \mathbf{p} \leq \mathbf{0} \quad (4.29)$$

$$\mathbf{H}\mathbf{x}_{t+1} - \mathbf{q} \leq \mathbf{y}_t - \mathbf{A}\mathbf{u}_t \quad (4.30)$$

$$-\mathbf{H}\mathbf{x}_{t+1} - \mathbf{q} \leq -\mathbf{y}_t + \mathbf{A}\mathbf{u}_t \quad (4.31)$$

for $t = 1, \dots, T$. Both \mathbf{C} and \mathbf{b} can be expressed by (4.20) as $\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$ and

$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$ where

$$\mathbf{b}_1 = \mathbf{0}_{(2T,1)} \quad (4.32)$$

$$\mathbf{b}_2 = \begin{bmatrix} \mathbf{y}_1 - \mathbf{A}\mathbf{u}_1 \\ -\mathbf{y}_1 + \mathbf{A}\mathbf{u}_1 \\ \vdots \\ \mathbf{y}_T - \mathbf{A}\mathbf{u}_T \\ -\mathbf{y}_T + \mathbf{A}\mathbf{u}_T \end{bmatrix} \quad (4.33)$$

$$\mathbf{C}_{11} = \begin{bmatrix} \mathbf{I}_{(r)} & -\mathbf{F} & \mathbf{0}_{(r,r)} & \cdots & \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} \\ -\mathbf{I}_{(r)} & \mathbf{F} & \mathbf{0}_{(r,r)} & \cdots & \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} & \cdots & \mathbf{0}_{(r,r)} & \mathbf{I}_{(r)} & -\mathbf{F} \\ \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} & \mathbf{0}_{(r,r)} & \cdots & \mathbf{0}_{(r,r)} & -\mathbf{I}_{(r)} & +\mathbf{F} \end{bmatrix} \quad (4.34)$$

$$\mathbf{C}_{12} = \mathbf{1}_{(2T)} \otimes [-\mathbf{I}_{(r)}, \mathbf{0}_{(r,n)}] \quad (4.35)$$

$$\mathbf{C}_{21} = \begin{bmatrix} \mathbf{H} & \cdots & \mathbf{0}_{(n,r)} & \mathbf{0}_{(n,r)} & \mathbf{0}_{(n,r)} \\ -\mathbf{H} & \cdots & \mathbf{0}_{(n,r)} & \mathbf{0}_{(n,r)} & \mathbf{0}_{(n,r)} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}_{(n,r)} & \cdots & \mathbf{0}_{(n,r)} & \mathbf{H} & \mathbf{0}_{(n,r)} \\ \mathbf{0}_{(n,r)} & \cdots & \mathbf{0}_{(n,r)} & -\mathbf{H} & \mathbf{0}_{(n,r)} \end{bmatrix} \quad (4.36)$$

$$\mathbf{C}_{22} = \mathbf{1}_{(2T)} \otimes [-\mathbf{0}_{(n,r)}, -\mathbf{I}_{(n)}] \quad (4.37)$$

Because vector \mathbf{X} has $((T + 2)r + n)$ entries, the matrix \mathbf{C} has $(2T(n + r))$ rows and $((T + 2)r + n)$ columns and the number of entries of the column vector \mathbf{b} equals to the number of rows of \mathbf{C} , i.e. $(2T(n + r))$. Under assumption that the SU model matrices are time-invariant (it is possible to specify the model more generally, see equations (1.19) and (1.20)), the construction of the \mathbf{b} and \mathbf{C} can be simplified into the form of (4.21) – (4.25).

Estimation of the Noise Bounds. Prior information on \mathbf{X} reflecting \mathcal{S}_0 and \mathcal{S}_2 is assumed to be in the form $\underline{\mathbf{X}} \leq \mathbf{X} \leq \overline{\mathbf{X}}$ using (4.27).

State and Parameter Estimation. The real-time (on-line) estimation provides the state and parameter estimates at each time step. A moving horizon estimator principle is used (see Subsection 3.2.2) and performs the Bayesian estimation on a sliding window of the length $\delta \in \mathbb{N}$ (it keeps the computational feasibility in the reasonable ranges and at the same time it allows to catch the slow parameter changes). Data $\{\mathbf{u}_{t-\delta}^T, \dots, \mathbf{u}_t^T, \mathbf{y}_{t-\delta}^T, \dots, \mathbf{y}_t^T\}_{t=\delta+1}^T$ and prior information on $\mathbf{x}_{t-\delta-1}$, \mathbf{p} and \mathbf{q} are used for an estimation of the states $\{\mathbf{x}_{t-\delta}^\top, \dots, \mathbf{x}_t^\top\}_{t=\delta+1}^T$.

Let's denote an estimated quantity as \mathbf{X}_t which has the following form as used in (4.5) for \mathbf{X}

$$\mathbf{X}_t = [\{\mathbf{x}_{t-\delta}^\top, \dots, \mathbf{x}_t^\top\}_{t=\delta+1}^T \quad \mathbf{p}^\top \quad \mathbf{q}^\top]^\top. \quad (4.38)$$

The superfluous state $\mathbf{x}_{t-\delta-1}$ and data item $\mathbf{y}_{t-\delta-1}$, $\mathbf{u}_{t-\delta-1}$ from the previous estimation step are integrated out from the posterior PDF in every time step t . This integration induces non-uniform term in the posterior PDF. This term is described by a piecewise polynomial function containing t powers of this state at each time instant t . With increasing t , the estimation becomes intractable because of increasing complexity of the support of the posterior PDF. An approximation of the non-uniform term in each step is applied - it consists in the replacing of the oldest state by its point estimate from the previous step. For more details see [60].

This technique can be applied also to the discrete-time state-space model with bounded innovations consisting of non-linear state and linear measurement equations. See [60] for necessary adjustments of the algorithm and reformulations of the MAP estimation into a form of a non-linear mathematical programming. In the following section, a generalization of this problem is introduced.

4.3 Non-Linear State-Space Model with Missing Data

A discrete-time state-space model defined by (1.22) and (1.23) is used for this chapter. Neither the state nor the measurement equation contain the exogenous

vector \mathbf{u}_t . Similarly to the previous sections, vectors of the state and output noises (\mathbf{v}_t and \mathbf{w}_t) are assumed to have uniform distribution on a multivariate box with the $\mathbf{0}$ center and unknown half-widths of the support intervals \mathbf{p} and \mathbf{q} , see specification (4.1).

In the following, only a single uninterrupted data outage is considered. This means that available data are now $\mathbf{y}\mathbf{y}_t^{a:b} \equiv \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_a, \mathbf{y}_b, \mathbf{y}_{b+1}, \dots, \mathbf{y}_t\}$ for $1 < a < b < t \leq T$. The number of all available data items is K , $K \leq T$. Note that an extension to the more data outages is straightforward.

The original PDF in a form of (1.38) is now replaced by

$$\begin{aligned} p(\mathbf{y}\mathbf{y}_T^{a:b}, \mathbf{x}\mathbf{x}_T, \theta) &\propto p(\mathbf{x}_0, \theta) \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{x}_t, \theta) p(\mathbf{x}_t | \mathbf{x}_{t-1}, \theta) \\ &= \left(\prod_{i=1}^r p_i \right)^{-T} \left(\prod_{j=1}^n q_j \right)^{-K} \mathbb{I}_{\mathcal{S}}, \end{aligned} \quad (4.39)$$

where T and K are numbers of available states and outputs, respectively, $\mathbb{I}_{\mathcal{S}}$ is the indicator of the support \mathcal{S} of this PDF. The set \mathcal{S} is a set of \mathbf{X} , see (4.5), such that (for given realization of $\mathbf{y}\mathbf{y}_t^{a:b}$) the noise terms in state and measurement equations of the non-linear state-space model are inside multivariate box defined by their uniform distribution (4.1) and restrictions (4.3) and (4.4)

$$\mathcal{S} = \{\mathbf{X} \in \mathcal{S}_3; \forall t \in \{1, \dots, T\} : |\mathbf{x}_{t+1} - \mathbf{g}(\mathbf{x}_t)| \leq \mathbf{p}, |\mathbf{y}_k - \mathbf{f}(\mathbf{x}_k)| \leq \mathbf{q}\}, \quad (4.40)$$

where \mathcal{S}_3 is the set of \mathbf{X} that meet the restrictions (4.3) and (4.4). The output outages are introduced by using different time indexes, t and k , in (1.22) and (1.23). Thus the measurement equation (1.23) is rewritten as

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{w}_k. \quad (4.41)$$

4.4 Non-Linear State and Noise Boundary Estimation

We aim to estimate vector \mathbf{X} considering the system described by the SU model (4.39). Again as in case of the linear system, a maximum a posteriori (MAP) estimation is used. It provides a point estimate $\hat{\mathbf{X}}$ of the internal quantity \mathbf{X} . The problem is equivalently transformed into an estimation of negative logarithm of a posteriori PDF in the following form

$$\hat{\mathbf{X}} = \operatorname{argmin}_{\mathbf{X} \in \mathcal{S}} \left(\sum_{i=1}^r \log(p_i) + \frac{K}{T} \sum_{j=1}^n \log(q_j) \right), \quad (4.42)$$

where \mathcal{S} is given by (4.40).

The previous implies that the problem of missing measurement data can be easily incorporated into the estimation algorithm. If all measurements are available, then $K = T$. The missing measurement causes that the corresponding entry of the output inequality coming from (4.41) is missing in the respective time instant. The more missing data, the smaller is the number of corresponding

constraint conditions in the set \mathcal{S} in (4.40) and the smaller is the weight of the second term in (4.42).²

The MAP estimate of (4.42) can be solved by the following non-linear programming form³

$$\min_{\mathbf{x} \in \mathcal{S}} \sum_{i=1}^r \log(q_i) + \frac{K}{T} \sum_{j=1}^n \log(p_j) \quad (4.43)$$

such that

$$\mathbf{c}(\mathbf{X}) \leq \mathbf{0}, \quad (4.44)$$

$$\underline{\mathbf{X}} \leq \mathbf{X} \leq \overline{\mathbf{X}}, \quad (4.45)$$

where

- the optimized function is a real function of \mathbf{X} , $\mathbb{R}^{n+(r+2)T} \rightarrow \mathbb{R}$,
- $\mathbf{c}(\cdot)$ is a real vector function that corresponds to the inequalities describing \mathcal{S} as (4.40)

$$\begin{array}{ll} \mathbf{x}_1 - \mathbf{f}(\mathbf{x}_0) - \mathbf{p} \leq \mathbf{0} & \mathbf{y}_1 - \mathbf{h}(\mathbf{x}_1) - \mathbf{q} \leq \mathbf{0} \\ -\mathbf{x}_1 + \mathbf{f}(\mathbf{x}_0) - \mathbf{p} \leq \mathbf{0} & -\mathbf{y}_1 + \mathbf{h}(\mathbf{x}_1) - \mathbf{q} \leq \mathbf{0} \\ \vdots & \vdots \\ \mathbf{x}_T - \mathbf{f}(\mathbf{x}_{T-1}) - \mathbf{p} \leq \mathbf{0} & \mathbf{y}_T - \mathbf{h}(\mathbf{x}_T) - \mathbf{q} \leq \mathbf{0} \\ -\mathbf{x}_T + \mathbf{f}(\mathbf{x}_{T-1}) - \mathbf{p} \leq \mathbf{0} & -\mathbf{y}_T + \mathbf{h}(\mathbf{x}_T) - \mathbf{q} \leq \mathbf{0} \end{array} \quad (4.46)$$

- $\underline{\mathbf{X}}$ and $\overline{\mathbf{X}}$ are known vectors - they stem from the set \mathcal{S}_3 , defined by inequalities in (4.3) and (4.4), and have the following form

$$\underline{\mathbf{X}} = \begin{bmatrix} \underline{\mathbf{x}}_0 \\ \mathbf{1}_{(T)} \otimes \underline{\mathbf{x}} \\ \mathbf{0}_{(r,1)} \\ \mathbf{0}_{(n,1)} \end{bmatrix} \quad \overline{\mathbf{X}} = \begin{bmatrix} \overline{\mathbf{x}}_0 \\ \mathbf{1}_{(T)} \otimes \overline{\mathbf{x}} \\ \overline{\mathbf{p}} \\ \overline{\mathbf{q}} \end{bmatrix}. \quad (4.47)$$

²To prevent the state-estimates divergence in case of measurement outage, data have to be present both at the beginning and end of estimated time interval.

³To solve the non-linear programming problem, the function *fmincon* from optimization toolbox of Matlab is used. This function starts with searching of $\hat{\mathbf{X}}$ at user supplied point $\hat{\mathbf{X}}_0$. The starting point $\hat{\mathbf{X}}_0$ of the optimization has to be set appropriately, because its improper setting causes numerical instability.

5. Particle Filters

Particle filters form another group of algorithms for incorporating linear and non-linear state constraints into filtration. Bayesian estimation of a dynamic process through Monte Carlo sampling tries to avoid problems of linear approximation and divergence, deterministic choice of sigma points or a fixed shape of distribution (either Gaussian for the moving horizon estimate or uniform for the SU model). The particle filters are more rigorous, but the trade-off is that they can require a lot of computational effort. Therefore they may not be feasible, depending on the application.

Roughly speaking, a Monte Carlo technique is a kind of stochastic sampling approach aiming to tackle the complex systems which are analytically intractable. The power of Monte Carlo methods is that they can attack difficult numerical integration problems. The attention of this chapter is focused on the Monte Carlo methods and particularly sequential Monte Carlo estimation. One of the attractive features of sequential Monte Carlo approaches lies in the fact that they allow on-line estimation by combining the powerful Monte Carlo sampling methods with Bayesian inference, at an expense of reasonable computational cost. In particular, the sequential Monte Carlo approach has been used in parameter estimation and state estimation, for the latter of which it is sometimes called particle filter. The basic idea of particle filter is to use a number of independent random variables called particles, sampled directly from the state space, to represent the posterior probability, and update the posterior by involving the new observations.

The particle filters operate by propagating many state estimates, particles, that are distributed according to the PDF of the true state. Just as the UKF can be considered as a generalization of the EKF, the particle filter can be considered as a generalization of the UKF. Given enough particles, a particle filter always performs better than a UKF, but this might be at the expense of unacceptable computational requirements, because the Monte Carlo methods are based on simulation, see [73].

State-constrained particle filtering has been solved by various methods. Some of these approaches can be used with Kalman filtering, such as reparameterizing the problem, see [1]. Other approaches are specific to particle filtering, such as modifying the particles' likelihood functions based on their level of constraint satisfaction, see [49] and [50], or generating process noise which ensures that the propagated particles satisfy the constraints, see [10]. Also, many of the methods can be potentially applied to constrained particle filtering, such as projection or PDF truncation, see Section 2.4. These methods could be applied to individual particles or they could be applied only to the state estimate at each time, giving rise to a large family of constrained particle filters.

In [51], inequality constraints are imposed by accept/reject steps in the algorithm. The appropriate constrained prior distribution is truncated or modified to satisfy the constraints, which ensures that the posterior also satisfies the constraints.

The approach introduced here relies on sequential Monte Carlo sampling to obtain the Bayesian solution in a computationally efficient manner without relying on simplifying assumptions. Given information about the state and measurement

equations and their parameters, the approach only needs to select the number of samples to simulate at each time point. This approach allows the distributions to adopt any shape at each time point, making the estimates quite accurate. This method is introduced in more details here.

This chapter is based on [6] and [51] if it is not stated otherwise. The original work, which introduced the Monte Carlo sampling for unconstrained estimation in dynamic systems, is in [15].

5.1 Sequential Importance Sampling

Using Monte Carlo sampling to solve estimation problems requires an approach for generating samples from the posterior, since direct drawing from the posterior distribution is not feasible. This may be accomplished via sequential importance sampling.

The sequential importance sampling (SIS) algorithm is a MC method that forms the basis for most sequential MC filters.¹ It is a technique for implementing a recursive Bayesian filter by MC simulations. The key idea is to represent the required PDF by a set of random samples with associated weights and to compute estimates based on these samples and weights. As the number of samples becomes large, this MC characterization becomes an equivalent representation to the usual functional description of the posterior PDF, and the SIS filter approaches the optimal Bayesian estimate.

One can draw samples $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N\}$, $N \in \mathbb{N}$, from a convenient importance function $\pi(\mathbf{x})$. A basic requirement on the importance function is that its support should include the support of the true distribution and the samples are easily generated. The estimate of the mean of a nonlinear function, $\mathbb{E} \varphi(\mathbf{x})$, regarding a posterior distribution $p(\mathbf{x})$ is given by²

$$\mathbb{E} \varphi(\mathbf{x}) = \int \varphi(\mathbf{x}) p(\mathbf{x}) \, d\mathbf{x} = \int \varphi(\mathbf{x}) \frac{p(\mathbf{x})}{\pi(\mathbf{x})} \pi(\mathbf{x}) \, d\mathbf{x} \approx \frac{1}{N} \sum_{i=1}^N \varphi(\mathbf{x}^i) \dot{q}^i, \quad (5.1)$$

where

$$\dot{q}^i = \frac{p(\mathbf{x}^i)}{\pi(\mathbf{x}^i)}. \quad (5.2)$$

Since it is usually impossible to sample from the true posterior, it is common to sample an easy-to-implement distribution, the importance density (also called a proposal density, see [16]).

Both $p(\mathbf{x})$ and $\pi(\mathbf{x})$ are assumed to be known up to a constant, in which case \dot{q}^i are normalized ($\sum_{i=1}^N \dot{q}^i = 1$),

$$\mathbb{E} \varphi(\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^N \varphi(\mathbf{x}^i) \dot{q}^i, \quad q^i = \frac{\dot{q}^i}{\sum_{i=1}^N \dot{q}^i} \quad (5.3)$$

¹The sequential MC approach is known variously as bootstrap filtering [30], the condensation algorithm [54], interacting particle approximations [59], or survival of the fittest [41].

²To avoid intractable integration in the Bayesian statistics, the PDF is empirically represented by a weighted sum of N samples drawn from the posterior distribution. When N is sufficiently large it is possible to approximate the true posterior (by the law of large numbers, as the number of samples goes to infinity, this estimate approaches the true value).

for $i = 1, \dots, N$.

Information about the relevant posterior distribution is contained in the pairs of samples and weights, $\{\mathbf{x}^i, q^i\}_{i=1}^N$, known as particles.

For a dynamic system, a prediction step is applied recursively by passing each sample through the state equation (1.13) to obtain samples corresponding to the prior at time t , $p(\mathbf{x}_t | \mathbf{y}_1, \dots, \mathbf{y}_{t-1})$. This prediction step utilizes information about process dynamics and model accuracy without making any assumptions about the nature of the dynamics and shape or any other characteristic of the distributions. Once the measurement \mathbf{y}_t is available, it can be used to recursively update the previous weights by the following equation, see [6]

$$\dot{q}_t^i \propto \dot{q}_{t-1}^i \frac{p(\mathbf{y}_t | \mathbf{x}_t^i) p(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i)}{\pi(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i, \mathbf{y}_t)}. \quad (5.4)$$

for $t = 1, \dots, T$.

The previous relation is based on the following approach. If the samples $\mathbf{xx}_t^i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_t^i\}$ were drawn from an importance density $\pi(\mathbf{xx}_t^i | \mathbf{yy}_t^i)$, where $\mathbf{yy}_t^i = \{\mathbf{y}_1^i, \mathbf{y}_2^i, \dots, \mathbf{y}_t^i\}$, then the weights are defined to be

$$q_t^i = \frac{p(\mathbf{xx}_t^i | \mathbf{yy}_t^i)}{\pi(\mathbf{xx}_t^i | \mathbf{yy}_t^i)}. \quad (5.5)$$

Returning to the sequential case, at each iteration, one could have samples constituting an approximation to $p(\mathbf{xx}_{t-1} | \mathbf{yy}_{t-1})$ and want to approximate $p(\mathbf{xx}_t | \mathbf{yy}_t)$ with a new set of samples. If the importance density is chosen to factorize such that

$$\pi(\mathbf{xx}_t^i | \mathbf{yy}_t^i) = \pi(\mathbf{x}_t | \mathbf{xx}_{t-1}, \mathbf{yy}_t) \pi(\mathbf{xx}_{t-1} | \mathbf{yy}_{t-1}) \quad (5.6)$$

then one can obtain samples $\mathbf{xx}_t^i \sim \pi(\mathbf{xx}_t | \mathbf{yy}_t)$ by augmenting each of the existing samples $\mathbf{xx}_{t-1}^i \sim \pi(\mathbf{xx}_{t-1} | \mathbf{yy}_{t-1})$ with the new state $\mathbf{x}_t^i \sim \pi(\mathbf{x}_t | \mathbf{xx}_{t-1}, \mathbf{yy}_t)$. To derive the weight update equation, $p(\mathbf{xx}_t | \mathbf{yy}_t)$ is first expressed in terms of $p(\mathbf{xx}_{t-1} | \mathbf{yy}_{t-1})$, $p(\mathbf{y}_t | \mathbf{x}_t)$ and $p(\mathbf{x}_t | \mathbf{x}_{t-1})$. Bayes' rule gives

$$\begin{aligned} p(\mathbf{xx}_t | \mathbf{yy}_t) &= \frac{p(\mathbf{y}_t | \mathbf{xx}_t, \mathbf{yy}_{t-1}) p(\mathbf{xx}_t | \mathbf{yy}_{t-1})}{p(\mathbf{y}_t | \mathbf{yy}_{t-1})} \\ &= \frac{p(\mathbf{y}_t | \mathbf{xx}_t, \mathbf{yy}_{t-1}) p(\mathbf{x}_t | \mathbf{xx}_{t-1}, \mathbf{yy}_{t-1})}{p(\mathbf{y}_t | \mathbf{yy}_{t-1})} p(\mathbf{xx}_{t-1} | \mathbf{yy}_{t-1}) \\ &= \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{t-1})}{p(\mathbf{y}_t | \mathbf{yy}_{t-1})} p(\mathbf{xx}_{t-1} | \mathbf{yy}_{t-1}) \\ &\propto p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{yy}_t - 1 | \mathbf{yy}_{t-1}). \end{aligned} \quad (5.7)$$

By substituting (5.6) and the last previous equation into (5.5), the weight update equation is

$$\dot{q}_t^i \propto \frac{p(\mathbf{y}_t | \mathbf{x}_t^i) p(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i) p(\mathbf{xx}_{t-1}^i | \mathbf{yy}_{t-1})}{\pi(\mathbf{x}_t^i | \mathbf{xx}_{t-1}, \mathbf{yy}_t) \pi(\mathbf{xx}_{t-1}^i | \mathbf{yy}_{t-1})} = \frac{p(\mathbf{y}_t | \mathbf{x}_t^i) p(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i)}{\pi(\mathbf{x}_t^i | \mathbf{xx}_{t-1}, \mathbf{yy}_t)}. \quad (5.8)$$

Furthermore, if $\pi(\mathbf{x}_t | \mathbf{xx}_{t-1}, \mathbf{yy}_t) = \pi(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_t)$ then the importance density becomes dependent only on \mathbf{x}_{t-1} and \mathbf{y}_t . This is particularly useful in the common

case when only a filtered estimate of $p(\mathbf{x}_t | \mathbf{y}_t)$ is required at each time step. From this point on, we will assume such a case, except when explicitly stated otherwise. In such scenarios, only \mathbf{x}_t^i need be stored, therefore one can discard the path \mathbf{x}_{t-1}^i and history of observations \mathbf{y}_{t-1} . The modified weight is then

$$\hat{q}_t^i \propto \hat{q}_{t-1}^i \frac{p(\mathbf{y}_t | \mathbf{x}_t^i) p(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i)}{\pi(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i, \mathbf{y}_t)}, \quad (5.9)$$

which is (5.4).

The updating step utilizes the measurement equation and information about the measurement error. Again, no assumptions about the type of model or distributions are required. The result of these prediction and updating steps is the particles at time t , $\{\mathbf{x}_t^i, \hat{q}_t^i\}_{i=1}^T$. Any posterior moment may then be calculated by equation (5.3). The resulting algorithm is fully recursive and computationally efficient since it avoids integration for obtaining the moments at each time step, nonlinear optimization in a moving window, or restrictive assumptions about the nature of the error or prior distributions and models.

The fundamental concept of sequential importance sampling is illustrated in Figure 5.1.

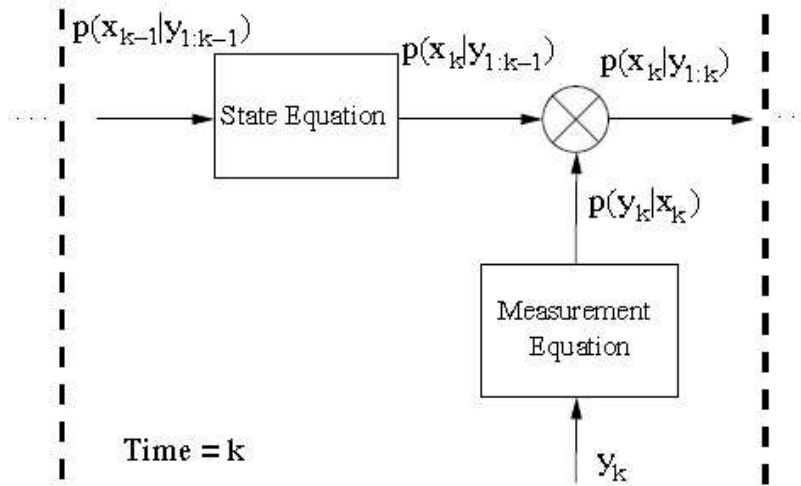


Figure 5.1: Sequential Importance Sampling Algorithm (Source: [14])

Convergence. Based on the previous calculation, convergence properties of the sequential Monte Carlo simulations are, see [48]

- the empirical distribution of the particles converges to the underlying true posterior density when the number of particles is increased,
- the asymptotic variance of the MC estimate stays bounded.³

Degeneracy Problem. A common problem with the SIS particle filter is a degeneracy phenomenon, where after a few iterations, all but one particle will

³The central limit theorem shows that under very weak conditions, the sequential MC approximation of the estimate based on the empirical distribution converges to the true estimate.

have negligible weight.⁴ The variance of the importance weights is increasing over time, which implies that a large computational effort is devoted to updating particles whose contribution to the approximation of the PDF is almost zero. This undesirable effect in particle filters can be solved by two methods, see [6]

- proper choice of importance density - additionally, the importance density requires an ability to sample from it, or
- resampling - it is used whenever a significant degeneracy is identified.

Resampling. An intuitive solution for the degeneracy is to multiply the particles with high normalized importance weights, and discard the particles with low normalized importance weights. This can be done in the resampling step. To monitor how bad the weight degeneration is, we need a measure. A suggested measure for degeneracy can be the so-called effective sample size, introduced in [47]. However in practice, the true effective sample size is not available, thus its estimate is used, see [16].

When the effective sample size is below a predefined threshold, the resampling procedure is performed. The procedure was also used in the rejection control that combines the rejection method and importance sampling. The idea is following: when the effective sample size is lower than the threshold, then each sample is accepted with some given probability and all the accepted samples are given a new weight, and the rejected samples are restarted and rechecked at the all previously violated thresholds. It is obvious that this procedure is computational expensive as a number of MC simulations increases. For details see the code for the sequential importance sampling in the Appendix.

Other Particle Filters. The sequential importance sampling algorithm forms the basis for most particle filters. The various versions of particle filters proposed in the literature can be regarded as special cases of this general SIS algorithm. These special cases can be derived from the SIS algorithm by an appropriate choice of importance sampling density and/or modification of the resampling step. The most commonly considered ones are

- sampling importance resampling filter (bootstrap filter),
- auxiliary sampling importance resampling filter,
- regularized particle filter,
- likelihood particle filter.

For a description of the methods and further extensions, see [16]

5.2 Constrained Sequential Importance Sampling

The proposed approach extends previous algorithms to satisfy inequality constraints. Equality constraints may be imposed by including them in the state

⁴Another problem is that although the approach is asymptotically convergent, in practice, a poor initial guess can cause the convergence to be very slow.

or measurement equations. We impose only constraints on state in the form of (2.84): $\mathbf{D}\mathbf{x}_t \leq \mathbf{d}$.

The implementation of the algorithm uses a convenient choice of importance function as

$$\pi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) = p(\mathbf{x}_t|\mathbf{x}_{t-1}). \quad (5.10)$$

This choice simplifies the recursive weight calculation by (5.4) to the following form

$$\dot{q}_t^i \propto \dot{q}_{t-1}^i p(\mathbf{y}_t|\mathbf{x}_t^i). \quad (5.11)$$

Updating the prior with the current information then only requires the likelihood value.

Algorithm for Estimation by Constrained Sequential Importance Sampling. This approach is represented by the following steps. For every $t = 1, 2, \dots, T$ run this sequence

1. For samples $i = 1, 2, \dots, N$:
 - Until $\{\mathbf{x}_{t+1}^i, \mathbf{v}_{t+1}\}$ satisfy the imposed constraints
 - Draw sample \mathbf{v}_{t+1} from system noise,
 - Calculate \mathbf{x}_{t+1}^i using state equation,
 - Assign a weight \dot{q}_t^i to \mathbf{x}_{t+1}^i .
2. Normalize \dot{q}_t^i to find q_t^i .
3. Implement resampling if necessary.

Enforcement of constraints in the sequential importance sampling is implemented by the extra steps during point 1. in the algorithm. These accept/reject steps evaluate the samples \mathbf{v}_{t+1} and the corresponding \mathbf{x}_{t+1}^i generated by the prediction step, via the state equation. Only those samples of the generated $\{\mathbf{x}_{t+1}^i, \mathbf{v}_{t+1}\}$ that satisfy constraints are accepted. Note that, the noise distributions (including the prior distribution) may be subject to the constraints and therefore drawing samples from these distributions may require another accept/reject step whenever it is inconvenient to sample directly from the underlying distributions.

Convergence. The algorithm ensures that the accepted particles are truly generated from the correct transition densities under the constraints. It means that the accept/reject operation leads to particles that correctly represent the posterior distribution, which ensures validity of the unconstrained convergence theorems also for the constrained simulations.

6. Application: Economic Data Filtering

The aim of this chapter is to apply three of early mentioned algorithms to real economic data. After a simple semi-structural model is introduced, its state-space representation is proposed. Then the model is applied to Czech economy data and some particular results are discussed, including inequality constraints which are used later. The presented methods are the Kalman filter with a probability density function truncation, the state uniform model filtration and particle filters with sequential importance sampling. Results of these methods are mutually compared and discussed.

6.1 A Monetary Policy Semi-Structural Model

Now the simplest type of a model, a small closed economy defined by three key variables and three structural equations, is introduced. The complete model is composed of these three structural equations, some identities and autoregression (AR) processes.

A reduced-form New Keynesian model, that is calibrated to fit stylized facts of the Czech economy, is presented. It is a simplified version of a generic Quarterly Projection Model used in several central banks, see e.g. [8]. The objective of this class of models is to help decide on an appropriate level of the policy interest rate, given the inflation target and the current state of the economy. A closed economy version of the model is used for a simplicity.¹

The structural equations of the model are:

- Aggregate Demand (IS curve),
- Aggregate Supply Block (Phillips curves), and
- Monetary Policy Rule.

Aggregate Demand. The aggregate spending relationship corresponds to the closed economy version of the traditional IS curve and takes the form

$$l_y_gap_t = a_1 l_y_gap_{t-1} + a_2 (r_{t-1} - r_tnd_{t-1}) + shock_l_y_gap_t, \quad (6.1)$$

where $l_y_gap_t$ is the deviation of the log of output from its noninflationary level, the output gap; $(r_{t-1} - r_tnd_{t-1})$ is a lagged deviation (gap) of the long-term real interest rate, r_{t-1} , from its neutral (noninflationary) level, r_tnd_{t-1} . This gap represents a monetary policy stance. Variable $shock_l_y_gap_t$ is an aggregate demand shock. The aggregate demand shocks are governed by a normal distribution and have no serial correlation. The coefficients a_1 and a_2 capture the persistence of output and the impact of the lagged real interest rate gap, respectively.

¹Although the Czech economy is an open one, the closed economy model is able to satisfactorily capture its behavior and can be easily extended to a form of an open economy model.

Aggregate Supply Block. The supply block consists of a standard backward looking Phillips curve for the overall consumer price index (CPI) inflation

$$dl_cpi_t = b_1 dl_cpi_{t-1} + b_2 l_y_gap_t + shock_dl_cpi_t. \quad (6.2)$$

The inflation depends on past CPI inflation values, dl_cpi_{t-1} , and on the current value of the real marginal costs approximated by the output gap, $l_y_gap_t$. A way of thinking about parameter b_1 is the persistence of inflation - the more persistent inflation, the higher is the coefficient. The coefficient b_2 captures an influence of the gap in the real marginal costs on inflation (the slope of the Phillips curve) and also measures the sacrifice ratio – how much output will be lost in order to bring inflation down by 1 percentage point.

Monetary Policy Rule. The model is closed by a policy reaction function of the monetary authority. The rule is defined typically for a central bank implementing inflation targeting monetary policy. For simplicity, we take the three-month interest rate to be an instrument of monetary policy. The authority is assumed to respond to deviations of inflation from its target and to the output gap.² The last period policy stance affects the current policy stance allowing the authority to smooth interest rates by adjusting them gradually to the desired level implied by the deviations of inflation and output from equilibrium,

$$i_t = c_1 i_{t-1} + (1 - c_1) [(r_tnd_t + dl_cpi_t) + c_2 dl_cpi_t + c_3 l_y_gap_t] + shock_i_t, \quad (6.3)$$

where i_t is the domestic short-term nominal interest rate and $shock_i_t$ is a policy shock. The policy neutral rate is $(r_tnd_t + dl_cpi_t)$.

The model includes other identities and transformations:

- an adjusted Fisher equation to relate nominal and real interest rates,

$$r_t = i_t - dl_cpi_t, \quad (6.4)$$

- a real gross domestic product (GDP) identity, dividing the log of GDP, l_y_t , into the log of potential product (noninflationary) growth, $l_y_tnd_t$, and output gap, and identities for a calculation of the annualized quarter-to-quarter (QoQ ann.) growth (in a log approximation) of real GDP, dl_y_t , and potential product, $dl_y_tnd_t$,

$$l_y_t = l_y_tnd_t + l_y_gap_t \quad (6.5)$$

$$dl_y_t = 4 (l_y_t - l_y_{t-1}) \quad (6.6)$$

$$dl_y_tnd_t = 4 (l_y_tnd_t - l_y_tnd_{t-1}), \quad (6.7)$$

- a law of motion for potential product growth and the real interest rate trend captured by an AR(1) with parameters d_1 and e_1 , and shocks $shock_dl_y_tnd$ and $shock_r_tnd_t$, respectively,

$$dl_y_tnd_t = d_1 dl_y_tnd_{t-1} + shock_dl_y_tnd_t. \quad (6.8)$$

$$r_tnd_t = e_1 r_tnd_{t-1} + shock_r_tnd_t \quad (6.9)$$

²Inflation targeting regime is aimed at the expected inflation in one year horizon. In our framework, the monetary authority does its decisions based on the current inflation for simplicity and because of a relative stable inflation dynamics in the Czech Republic.

6.1.1 State-Space Representation of the Model

The model consists of a set of linear equations (6.1) – (6.8). All the model variables are declared with a time t subscript, $t = 1, 2, \dots, T$. These are state variables and a disturbance term (structural shocks and measurement errors),

$$\mathbf{x}_t = [l_y_gap_t, dl_cpi_t, i_t, r_t, l_y_t, dl_y_t, l_y_tnd_t, dl_y_tnd_t, r_tnd_t]^\top \quad (6.10)$$

$$\mathbf{v}_t = [shock_l_y_gap_t, shock_dl_cpi_t, shock_i_t, shock_dl_y_tnd_t, shock_r_tnd_t]^\top, \quad (6.11)$$

and the system can be rewritten as

$$\mathbf{B}\mathbf{x}_t = \mathbf{C}\mathbf{x}_{t-1} + \mathbf{D}\mathbf{v}_t, \quad (6.12)$$

where

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -b_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -(1-c_1)c_3 & -(1-c_1)(1+c_2) & 1 & 0 & 0 & 0 & 0 & 0 & -(1-c_1) & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.13)$$

$$\mathbf{C} = \begin{bmatrix} a_1 & 0 & 0 & a_2 & 0 & 0 & 0 & 0 & -a_2 & 0 \\ 0 & b_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & d_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & e_1 \end{bmatrix} \quad (6.14)$$

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.15)$$

The model is solved numerically using the Matlab software and IRIS Toolbox only.³ The model code in Iris is in Appendix 1.

³IRIS is a free, open-source toolbox for macroeconomic modeling and forecasting in Matlab. It integrates core modeling functions (such as flexible model file language, simulation, estimation, or forecasting) with a wide range of supporting features (such as time series analysis, data management, or reporting).

After a simple algebra assuming an existence of \mathbf{B}^{-1} , the state equations has the following form

$$\mathbf{x}_t = \mathbf{B}^{-1}\mathbf{C}\mathbf{x}_{t-1} + \mathbf{B}^{-1}\mathbf{D}\mathbf{v}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{B}^{-1}\mathbf{D}\mathbf{v}_t, \quad (6.16)$$

where $\mathbf{F} = \mathbf{B}^{-1}\mathbf{C}$. Notice that matrix $\mathbf{B}^{-1}\mathbf{D}$ introduces an additional connection among the structural shocks included in \mathbf{v}_t (\mathbf{v}_t can be respecified to enter the equation directly multiplied only by an identity matrix) and rescales a covariance matrix of the disturbance term (which is important for a calibration of standard deviations of the shocks, see later). The last equation is a state equation of the state-space representation of the model.

Compared to the 9-element state vector, the measurement vector and output disturbance (measurement error) vector consists of only 3 elements,

$$\mathbf{y}_t = [\text{obs_dl_cpi}_t, \text{obs_dl_yt}, \text{obs_it}]^\top \quad (6.17)$$

$$\mathbf{w}_t = [\text{meas_dl_cpi}_t, \text{meas_dl_yt}, \text{meas_it}]^\top. \quad (6.18)$$

Because there is a one-to-one mapping of the measurement variables to the state ones (common in economic practice) and there is no control variable \mathbf{u}_t (see the data description), the measurement equation has the following form

$$\mathbf{y}_t = \mathbf{A}\mathbf{u}_t + \mathbf{H}\mathbf{x}_t + \mathbf{w}_t = \mathbf{0}\mathbf{u}_t + \mathbf{H}\mathbf{x}_t + \mathbf{w}_t = \mathbf{H}\mathbf{x}_t + \mathbf{w}_t. \quad (6.19)$$

Then $\mathbf{A} = \mathbf{0}_{(3,1)}$ and \mathbf{H} is a (3×9) type matrix including only ones and zeros. Notice that the measurement equation contains also a measurement error term capturing an inaccuracy in observing and measuring economic variables.

6.1.2 Data

The model is applied to the Czech economy and the used data are the followings

- CPI: the time series is published by the Eurostat as a monthly index with a base in 2005, seasonally unadjusted.
- Real GDP: the time series is calculated from the nominal GDP adjusted by the GDP deflator, both published by the Eurostat. They are available on quarterly frequency and seasonally unadjusted as the nominal GDP level in millions of Czech crowns and the price index 2000=100 based on Czech crowns.
- Short-Term (Three-Month) Nominal Interbank Interest Rate: average annual reference interest rate with three-month maturity on the Czech interbank financial market, published by Eurostat on a monthly basis.

Time series of CPI and real GDP in $100 \log(\cdot)$ transformation are seasonally adjusted by Census X12 ARIMA (part of IRIS). The seasonal pattern, on the available range, is depicted in Figure 6.1. There are no missing data in the used sample.

A historical range for the filtration and parameter calibration was set for 2006Q1–2012Q4 period. At the beginning of 2006, the Czech National Bank

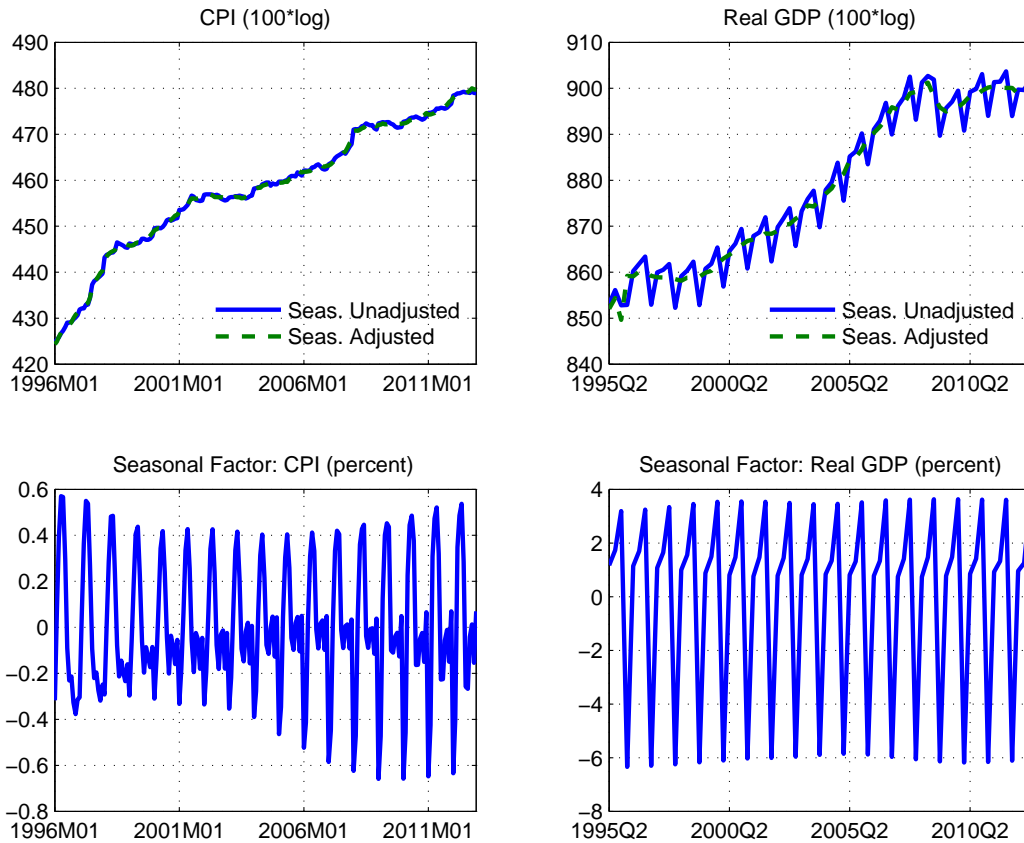


Figure 6.1: Seasonal Adjustment of Data (Source: author's calculation)

started to implement a point headline inflation targeting instead of a target band. Starting at this point ensures a consistency of the implemented monetary policy.

Because the model is the quarterly one, nominal interest rate and CPI are converted by a simple arithmetic average into a quarterly frequency. Except the interest rate, real GDP and CPI inflation are calculated as the natural logarithm and growth rates in QoQ annualized terms for these relevant variables. Subsequently, mean values of the variables are subtracted

- $obs_dl_cpi_t$: CPI inflation target announced by the Czech National Bank is taken away from the CPI inflation. The inflation target captures the mean value of inflation. It was set at 3 percent for 2006Q1–2009Q4 period and 2 percent afterwards,
- $obs_dl_y_t$: a mean value of 1.71 percent QoQ ann. (for 2006Q1–2012Q4) is subtracted from the QoQ ann. real GDP growth,
- obs_i_t : the calculation is done in the same way as for the real GDP growth, using a mean value of 2.16 percent p.a.

Subtracting the mean values and using the above introduced specifications of equations mean that the model has a gap form.⁴

⁴Using original data, including their mean values means, would require to extend the measurement equation by adding a control variable to capture the means.

Figure 6.2 shows both original data including its mean values and the time series used as a measurement variables for the model. Visual inspection of data is a useful check of your prior intuition about the economy.

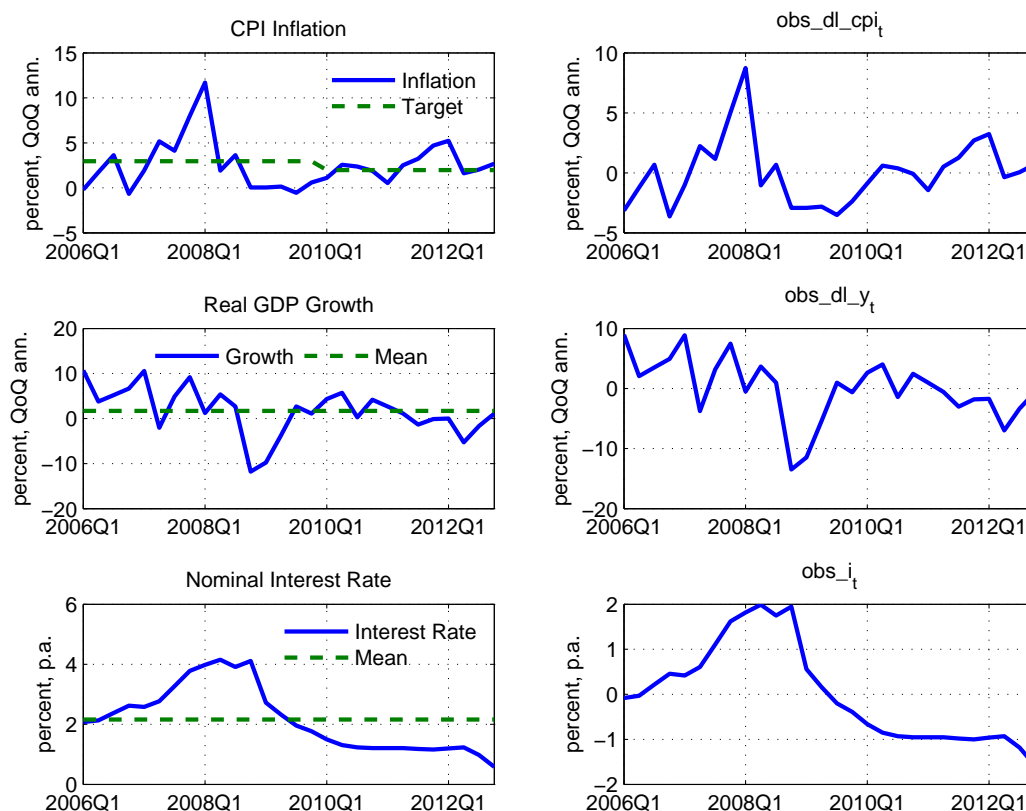


Figure 6.2: Measured Variables of Model (Source: author's calculation)

6.1.3 Economic Interpretation

Before we impose inequality constraints to some states, we discuss calibration and filtration with respect to their analysis and interpretation first.

In setting the parameters values we follow

- the economic theory,
- experience, and
- stylized facts about the economy, based on impulse responses, variance decomposition, historical simulations, etc.

The calibration of parameters and standard deviation of the structural shocks and measurement errors is presented in Table 6.1.

The KF estimates all the unobserved variables, gaps and trends, based on the observed variables and the model structure, ensuring that the results are model- and calibration-dependent. Due to the presence of trends within the model and resulting non-stationarity of some variables, the unconditional variance does not have a finite value and a diffuse KF is applied. In this subsection, a smoothing

a_1	0.60
a_2	-0.30
b_1	0.50
b_2	0.30
c_1	0.80
c_2	0.20
c_3	0.10
d_1	0.70
e_1	0.80
std: <i>shock_ly_gap</i>	1.40
std: <i>shock_dl_cpi</i>	1.50
std: <i>shock_i</i>	0.40
std: <i>shock_dl_y_tnd</i>	1.00
std: <i>shock_r_tnd</i>	0.70
std: <i>meas_dl_cpi</i>	0.75
std: <i>meas_dl_y</i>	0.75
std: <i>meas_i</i>	0.25

Table 6.1: Calibration of the Model

step of the KF is employed which, as opposed to the prediction and filtration steps, uses complete information from observed data.

A smoothed values of the measurement error for all three measurement variables reflect the model calibration and are depicted in Figure 6.3. CPI inflation without measurement error is smoother (reflecting a calibrated persistency for the Phillips curve). In case of real GDP growth and the nominal interest rate, measurement errors are almost negligible.

Although the KF algorithm itself appears to be quite general and straightforward, its successful implementation tends to be very problem-specific, relying heavily on judgments to adjust and tune process and the model. In case of the KF, these rules are recommended to follow, see [22]:

- Understand your data and processes generating data (maximum and minimum ranges, etc.) and learn by looking at available data and appreciating what kind of information is likely to be available to the filter. It can be quite pointless designing a filter without knowing what information will be available for use,
- Model your system because the performance of the filter will be directly dependent on the adequacy of this model. To build as accurate a *truth model* as possible means to describe all aspects of the process to be estimated. On the other hand, it is necessary to reduce the model to have a direct and significant impact on filter performance. This has to be done on a case-by-case basis,
- Coding, which is very often the easiest part of the implementation,
- The recursive formulation of the KF algorithm means that we must provide some reasonable *guess* for the initial conditions,

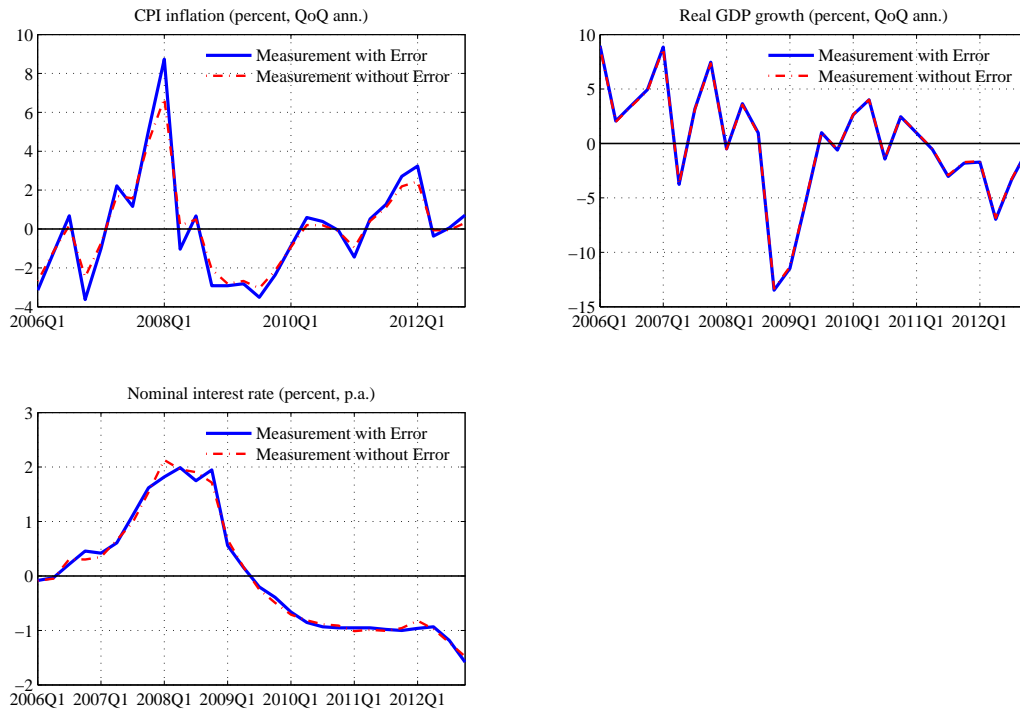


Figure 6.3: Measurement Errors after Filtration (Source: author's calculation)

- Run filtration and do analyses of results, which can be based on discrepancy sequence, steady-state performance, robustness of filtration, interpretation of results, etc.

The last item of the rules covers the rest of this subsection.

The filtered trend and gaps are presented in Figure 6.4. Despite the model simplicity and the fact that the Czech economy heavily depends on the world economy development (especially Eurozone which is not included in the model framework), potential growth of the economy is higher before the global economic crisis and slowdowns after 2008. Similarly, the filtration identifies positive output gap till the end of 2008 and negative (or neutral) afterwards. Additionally, after 2009 there is a clear *w-shape* of the output gap, which is unofficially discussed and believed among economists. After 2010, a neutral real interest rate gap corresponds to a fact that monetary policy is not efficient in influencing the real economy. The authority started to use additional non-ordinary instruments and monetary policy transmission channels (a set of these tools is usually labeled as a quantitative easing).

A decomposition of the structural equations for the output gap, CPI inflation and nominal interest rate into structural shocks (the upper row) and into a contribution of the structural variables (the lower row) is shown in Figure 6.5. Based on this set of pictures, it is possible to conclude that the structure of the model and its calibration approximate the Czech economy behavior satisfactorily and the model can be used for an analysis of the monetary policy implementation.

Based on the previous results, it is useful to discuss smoothed trajectory of the real interest rate trend and neutral nominal monetary policy rate in more details. Both variables with their mean values are presented in Figure 6.6 (trajectories are calculated out of the KF). In reality, it is possible that the real interest rate trend

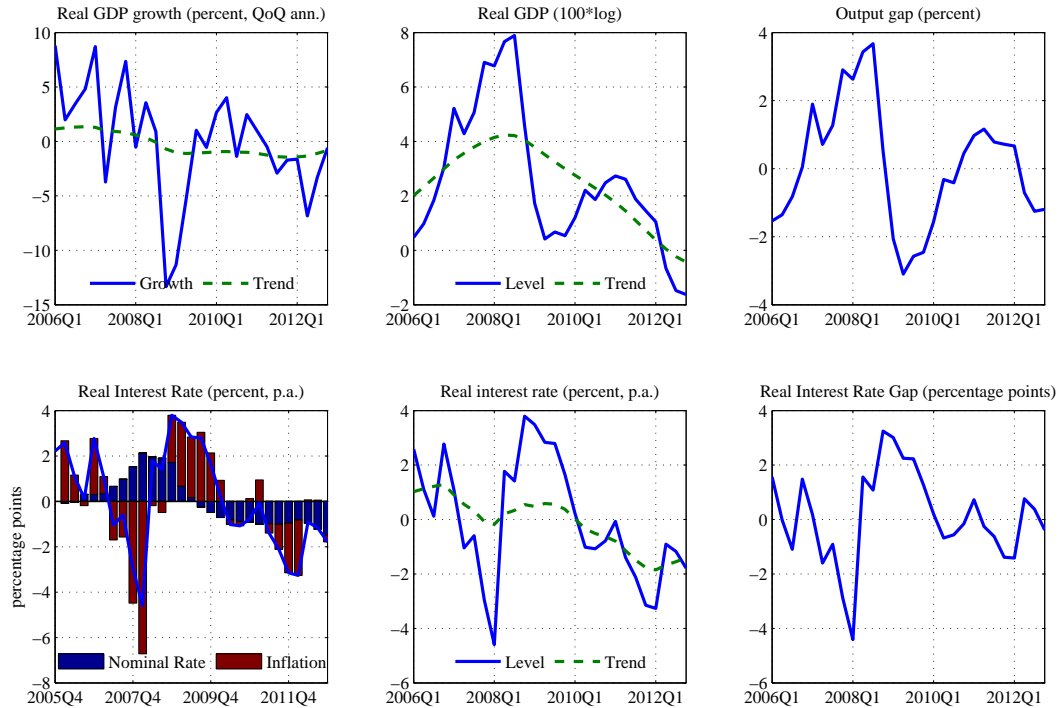


Figure 6.4: Trends after Filtration (Source: author's calculation)

can be negative temporarily, but not for a long period. The negative real interest rate trend implies dissaving and disinvestment behavior changing consumption dynamics, redistributes welfare from creditors to debtors, produces changes in securities trading, etc. Let's impose a constraint to the real interest rate trend to remain non-negative. This means that the demeaned real interest rate trend in the model, r_tnd_t , should not drop below -0.86 percent p.a. during the period of 3 percent inflation target and 0.16 percent p.a. after the inflation target was reduced to 2 percent.⁵ The negative rate appeared partly during 2008 (because of high CPI inflation fueled by a sharp increase in imported world food and oil prices) and especially after 2011 (the common monetary policy was not efficient in influencing a real economy performance as a result of the global crisis). For simplicity, let's constrain the real interest rate trend at nonnegative values for the whole range used for filtration.

In other words by imposing the constraint, we are asking about a stance of fiscal policy: in case there is a coordination between monetary and fiscal authorities, what kind of fiscal expenditures (measured by fiscal impulses and shocks to the output gap through the IS curve) is necessary to run that the real interest rate remain non-negative in the long run.

Similar problem is connected to the neutral monetary policy rate. This rate, capturing a trend in nominal interest rate in the model, should not be negative for a long time (the Czech National Bank can not set negative nominal rates as its equilibrium). This is not the case of the previous filtration, however, by raising the real interest rate trend, there will be an increase in the neutral rate (the

⁵The inflation target was set at 3 percent for 2006Q1-2009Q4 and 2 percent afterwards. This together with the mean value of 2.16 percent p.a. for the nominal interest rate implies the previously calculated numbers based on the Fisher equation, see data subsection and equation (6.4).

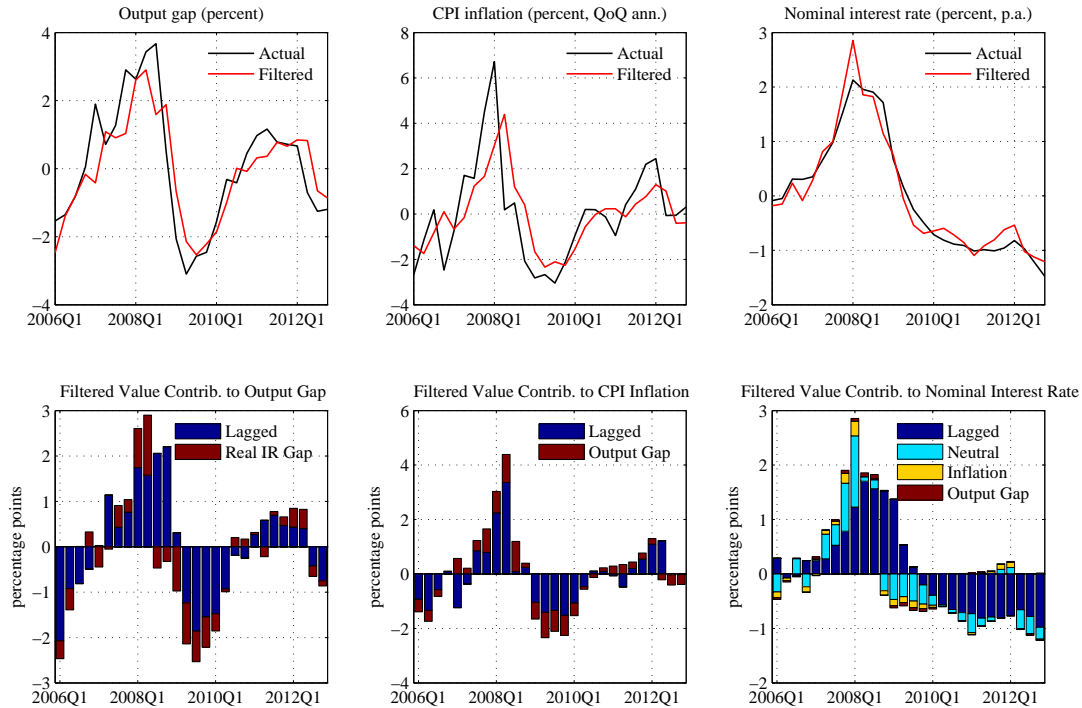


Figure 6.5: Structural Variables Decomposition after Filtration (Source: author's calculation)

neutral rate is a real interest rate trend plus CPI inflation identity). Reflecting this fact, any constraint for non-negativity of the neutral interest rate should be a passive one and thus it is not imposed.⁶

6.2 Imposing an Inequality Constraint

In this subsection, an inequality constraint for the real interest rate trend (as a state variable which is not directly measured) is imposed

$$r_tnd_t \geq 0, \quad (6.20)$$

for all t . This means that a state prediction is constrained in a form of an inequality, see (2.86).

For solving an inequality constrained state vector problem, it is possible to use a variety of methods introduced in the previous sections. The following three methods are applied to economic data and their results are compared

- Probability density function truncation,
- State uniform model filtration, and
- Sequential importance sampling.

A motivation behind this selection is following. The PDF truncation, using a KF calculation, is one of the most commonly employed method. The SU filtration with noise boundary estimation is used for some economic data for the first

⁶The problem of a zero-nominal-interest-rate bound appeared during the global crisis and still attracts a lot of attentions, see e.g. [80].

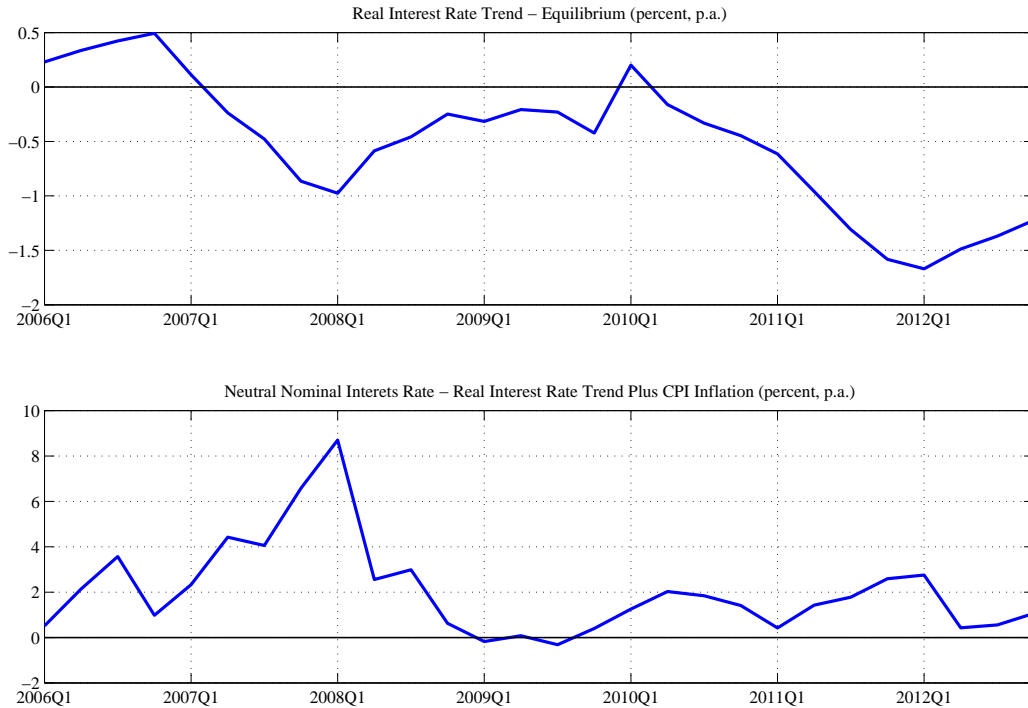


Figure 6.6: Neutral Nominal and Real Interest Rate Trend Filtration (Source: author’s calculation)

time here. The SIS algorithm, as one of the method based on Monte Carlo simulations, has recently started to be popular and formed an alternative against the *traditional* Kalman filtering.⁷ Additionally, it is possible to relax an assumption about Gaussian state estimation and after a simple adjustment, the SIS can be confronted with the SU filtration relying on uniform distribution of disturbances.

For a comparison, all three algorithms are run with the same data set (including the time span), model specification and calibration of structural parameters and standard deviations of the shock (in case of the SU model filtration, a bound for filtered values of the shocks is estimated) and also for the initial conditions. The initial conditions for all state variables are set to zero (their mean values). Specific settings of particular algorithms are discussed within the relevant parts.

Due to a comparison reason, the prediction step of algorithms is used and a comparison of the methods is based on a predicted error for the measurement equation (predicted measurement error).

The prediction step of the filtration at time t for $t + 1$ use all available information till time t , but not the full information set of observations till T compared to the smoother.⁸ Thus the results can suffer by a big error. See Figure 6.7 to demonstrate the difference between the trajectories of the measurement variables. Using the state-space model representation and calibration of the parameters introduced earlier, there is a comparison between smoothed and predicted trajectories. Whilst the difference for nominal interest rate is almost negligible, big

⁷A comparison among MC sampling, the extended Kalman filter and moving horizon estimation (another frequently used method) has been already covered by [14] and can be easily extended for a state constraint problem.

⁸As an alternative to the smoothing step in the Kalman filter, an off-line estimation in the SU model can be used.

errors are contributed for real GDP growth.

Similarly, there is a different dynamics for the real interest rate trend, see Figure 6.7. However, the problem of negative real interest rate trend appears after 2009 and it is valuable to continue with the analysis based on the prediction step of filtrations.

The Matlab codes for a core part of all three algorithms are supplied in Appendices.

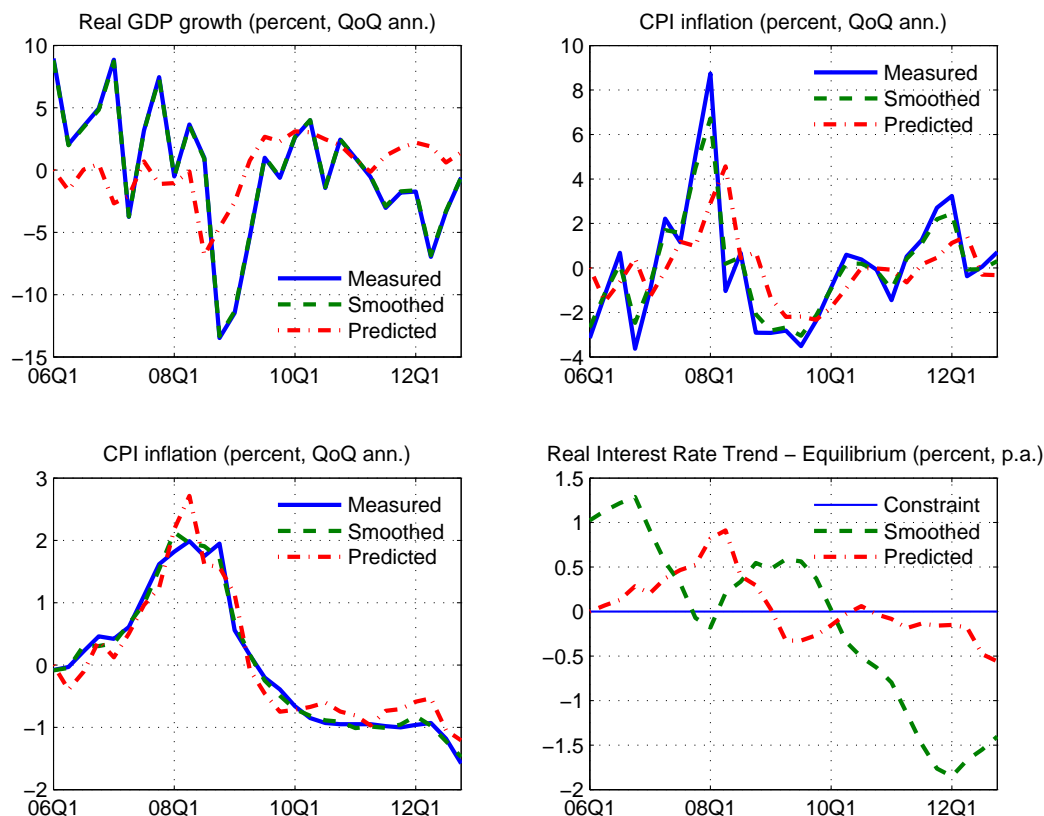


Figure 6.7: Kalman Filter: Smoothed and Predicted Values of Measurement Variables and Real Interest Rate Trend (Source: author’s calculation)

6.2.1 Probability Density Function Truncation

Results of the prediction step for the PDF truncation are depicted in Figure 6.8. The KF with the PDF truncation preserves relatively big prediction error for the real GDP growth measurement. Prediction step of the filtration for CPI inflation is smoother compared to its measurement.

By imposing the inequality constraint, the real interest rate trend is non-negative and hovers between 0.6 – 0.9 percent after 2009 (during the period when the unconstrained KF prediction of the trend sticks negative on average). This implies more negative real interest rate gap compared to the gap without any constraint (the predicted one by the unconstrained KF) after 2009. It contributes to a faster closing of the output gap during 2009 and more positive gap afterwards. In other words, the simulation shows that a fiscal policy expansion, supporting domestic aggregate demand, enables the monetary authority to keep non-negative

real interest rate (and consequently also non-negative neutral nominal interest rate which is not plotted in the figure).

Technically, a modified Gram-Schmidt orthonormalization calculation was used. The code is partly based on [73].⁹ It can be easily modified to implement an inequality constraint for an upper bound for any of the state variables or to use the constraint for an updating step of the KF - using inequality constraints in a form of (2.85).

6.2.2 State Uniform Model

Results of the SU model filtration exhibit relatively low prediction error for real GDP growth, similarly also for CPI inflation and nominal interest rate with an exception of some periods, see Figure 6.9. Real interest rate trend and potential growth are stick on zero values apart from the first two years of the historical data range. After this period, the dynamics of both output and real interest rate gap is similar to the unconstrained KF. This results into a higher real GDP growth and higher nominal interest rate (it is filtered through measurement errors at the end of the sample).

The filtration logically says that in case that the real interest rate trend can not be negative, it remains zero. It partially increases nominal interest rate, to eliminate an effect of monetary policy restriction on the output gap, and the fiscal policy must be more pro-active. It results in higher real GDP growth. The real interest rate trend, real GDP growth and CPI inflation show relatively high volatility in this simulation. Whilst the high volatility is a desirable characteristic of measurement variables in this simulation, it is something undesirable for trends. This property reflects disturbances with a uniform distribution. The estimated measurement error boundaries for measurement variables, see (4.4), are shown in Figure 6.10.

A calibration of the standard deviation of the structural shocks and measurement errors is not used, but it is estimated as a part of the filtration process. The algorithm is based on a linear programming and the Optimization Toolbox in Matlab is required. As it was described in Chapter 4, it is necessary to set lower and upper boundaries for the state vector, state and output noise, and a memory for the linear programming – length of a sliding window δ , see (4.38). By iterative simulations, the boundaries are finally set to give a reasonable results and the memory for programming is four. The coded algorithm was provided by the supervisor of the thesis.

6.2.3 Sequential Importance Sampling

The MC simulation was run for 2000 samples, however, there is no convergence of the chain. Despite no statistical tests, only a visual inspection of the convergence was carried out, but a rerunning of simulations has not resulted into the same results.¹⁰ Even without the convergence, the prediction of the measurement and

⁹See <http://academic.csuohio.edu/simond/ConstrKF>.

¹⁰The code for running the sequential importance sampling itself is not optimized, but whilst 500 simulations required 44 seconds, 2000 simulations required 2912 seconds, using Intel Core Duo CPU, P8600 @ 2.4 GHz, 2 GB RAM.

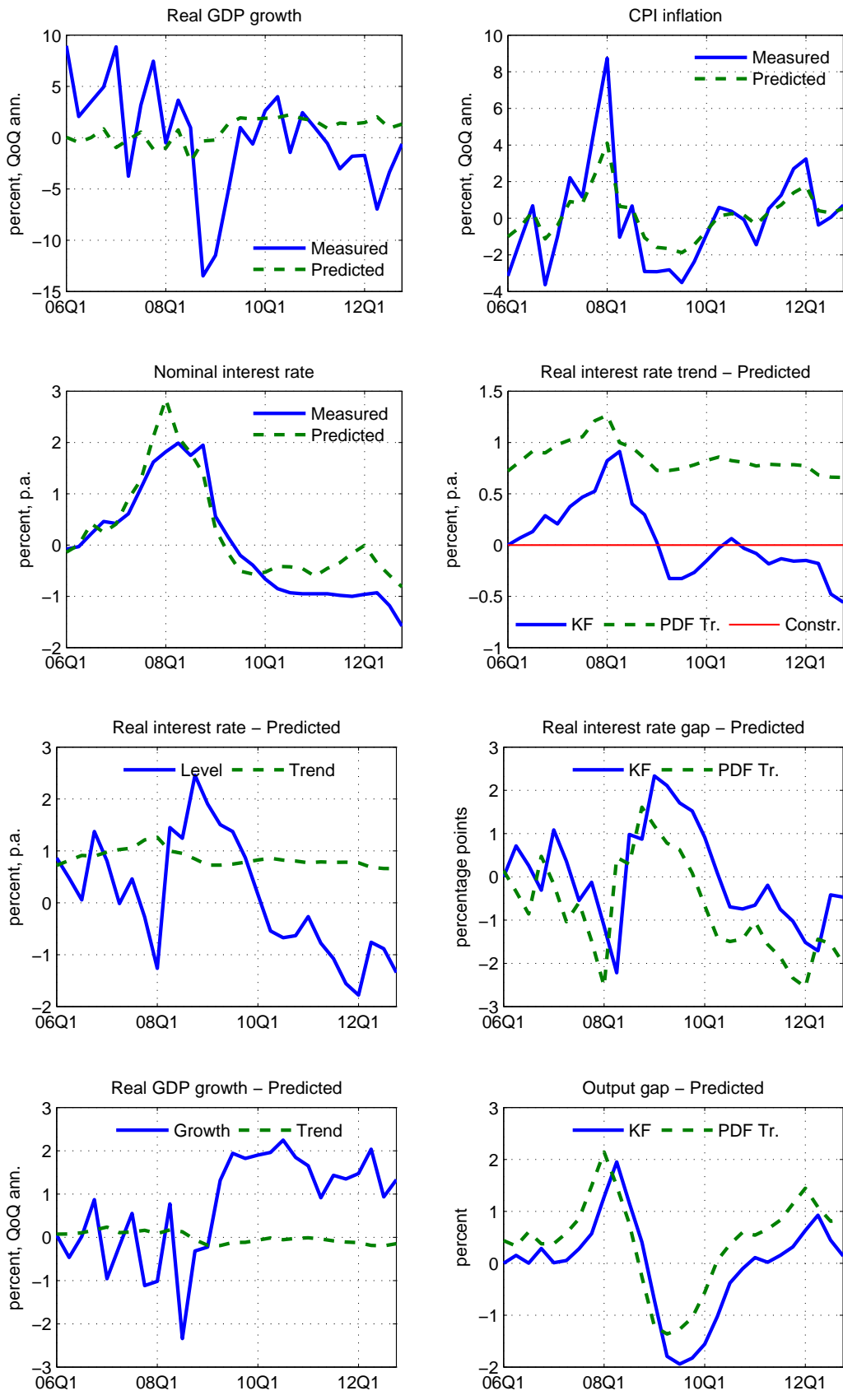


Figure 6.8: PDF Truncation: Results (Source: author's calculation)

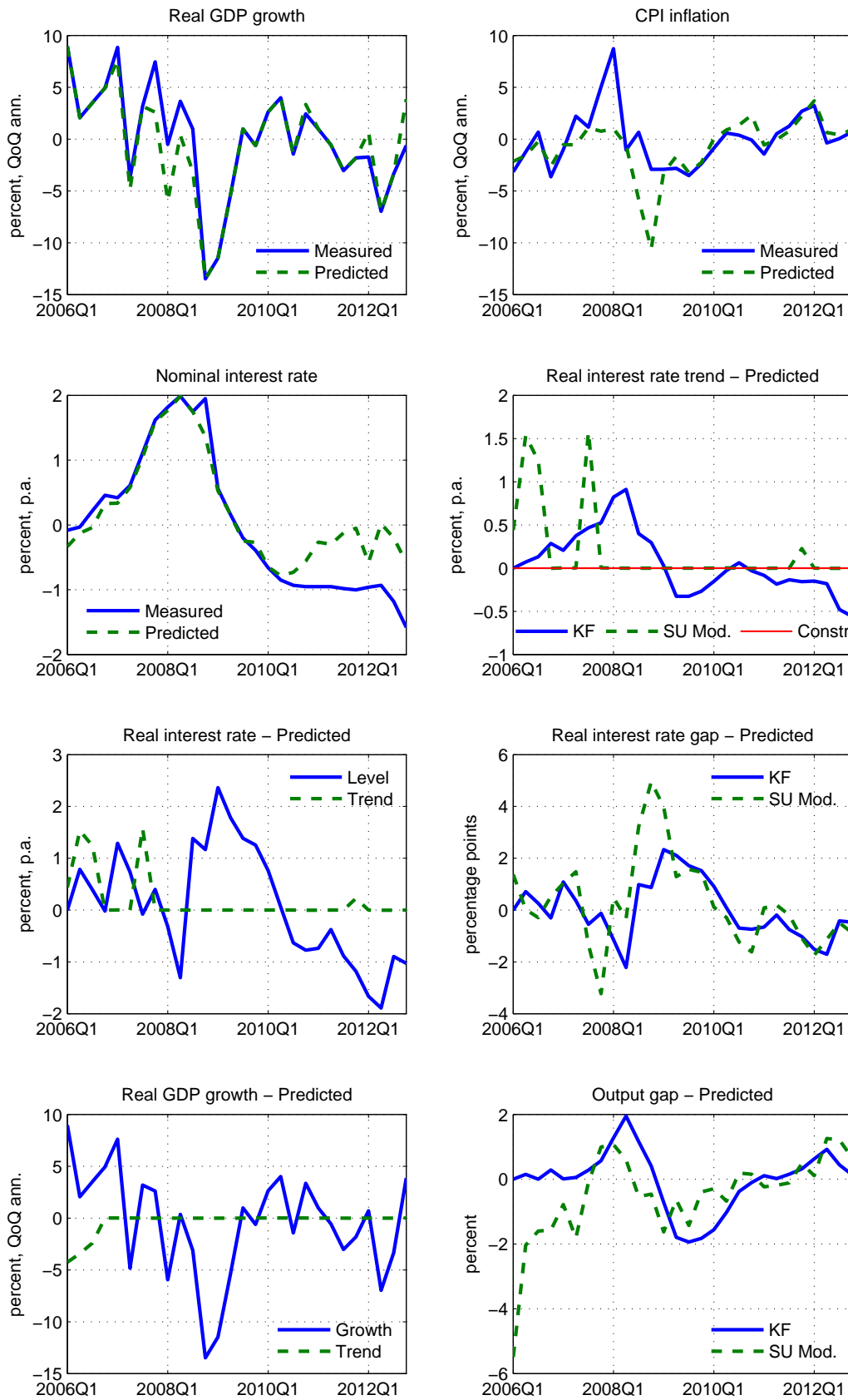


Figure 6.9: SU Model: Results (Source: author's calculation)

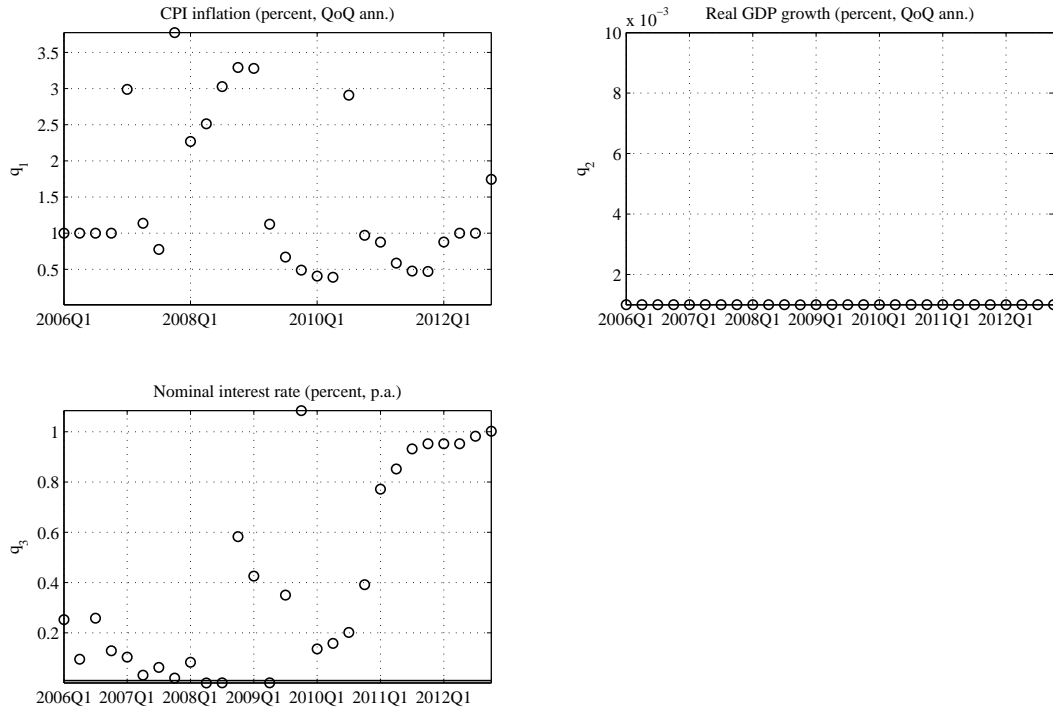


Figure 6.10: SU Model: Noise Boundaries for Measurement Variables (Source: author's calculation)

state variables does not differ significantly and keeps the same dynamics on average. Reflecting this non-convergence problem, the presented results are believed to be satisfactory for general discussion and a comparison of introduced methods in this section.

Predicted measurement variables closely follow the observed pattern with some exceptions particularly for real GDP growth, see Figure 6.11. A prediction of the real interest rate trend is relatively volatile and after 2011 implies a negative real interest rate gap for the following two years. This implies a positive output gap for the same period (it means positive aggregate demand shocks as fiscal policy impulses).

The code for a SIS calculation is marginally based on [14], where errors were fixed and a section for imposing inequality constraints was added. To run the simulations, it is necessary to specify a number of simulations and a threshold to avoid a degeneracy problem. The threshold was set as a third of the total number of MC simulations.

6.3 Comparison of Algorithms

In this part there is a comparison of the previously introduced and applied three algorithms with respect to their speed, accuracy, reliability and initiation.

To compare a performance of the algorithms, it can be used several criteria. Regarding [81], the accuracy of the state estimate can be quantified by the root mean square error (RMSE). Next, it can be assessed how informative (based on [52]) the state estimate is by evaluating the mean square error (MSE), in case of the Kalman filter the mean trace of $\mathbf{P}_{t|t}$ or $\mathbf{P}_{t|t-1}$ for $t = 1, 2, \dots, T$. It measures

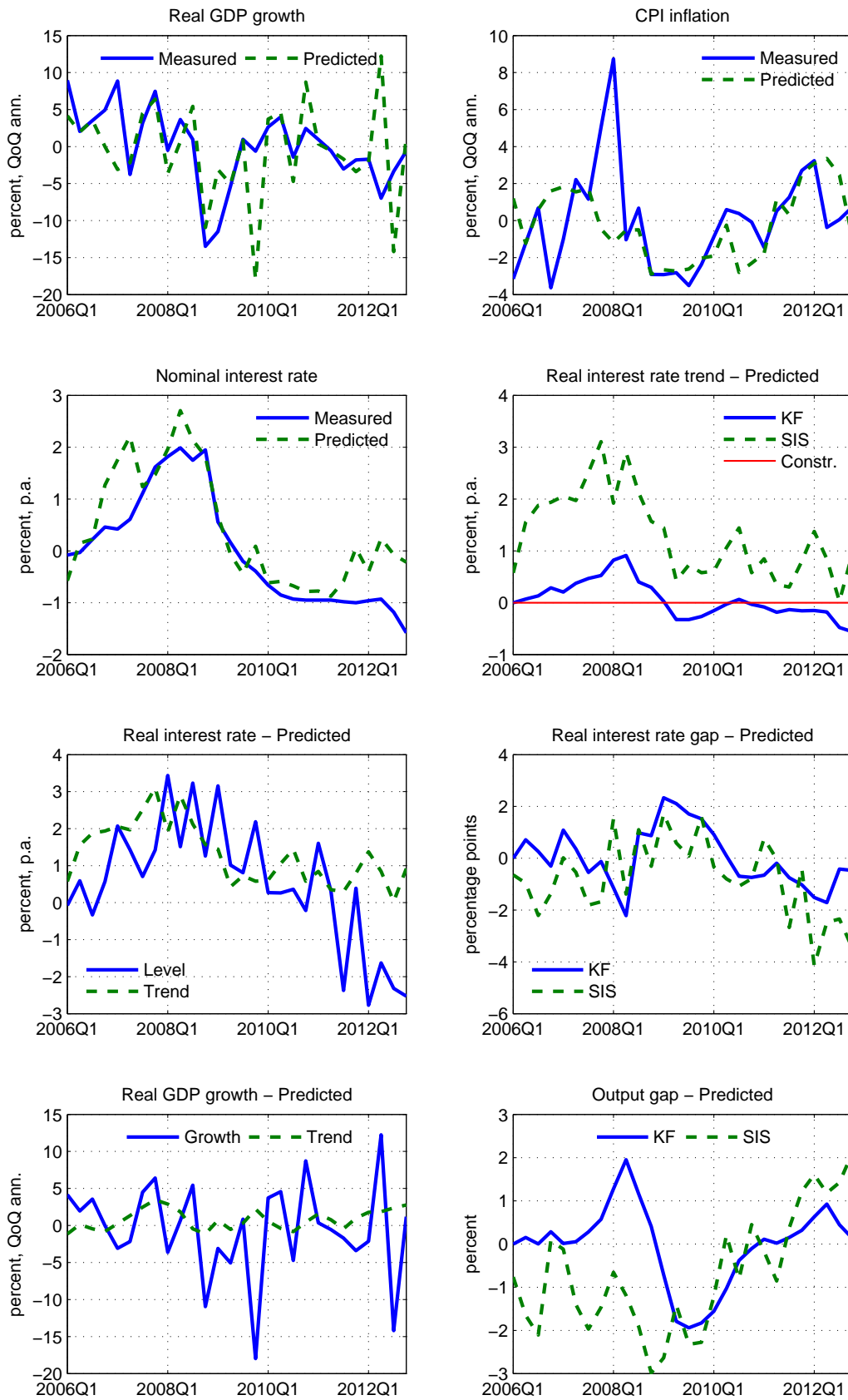


Figure 6.11: SIS: Results (Source: author's calculation)

the uncertainty of an estimate of a state. Another way is to use an absolute error estimate, see [60]. It is defined as a absolute difference between values of measured variables and estimated values. In case of running MC simulations, [51] suggests to evaluate the performance by the overall mean-squared error and specifically by the MSE averaged over realizations for each time. By examining this adjusted statistics over time, it is likely to indicate the long-term behavior of the tested method and provide insight into the distribution of errors over time.

Accuracy and Reliability. To compare the results of these heterogenous methods, it is necessary to use general and easily comparable and interpretable statistics - mean, standard deviation and root mean square error for the prediction error for the measurement variables (CPI inflation, real GDP growth and nominal interest rate). A prediction ability of the method is compared with observed data and then a predicted measurement error is calculated. The statistics are summarized in Table 6.2.

	CPI Inflation	Real GDP Growth	Nominal Interest Rate
Mean			
KF Unconstrained	-0.0313	-0.0271	0.0608
PDF Truncation	0.1242	0.7061	0.2590
SU Filtration	1.3934	0.3572	0.3411
PF - SIS	-0.2121	-0.5132	0.4040
Standard Deviation			
KF Unconstrained	2.4210	5.0085	0.3858
PDF Truncation	1.5045	5.5293	0.4117
SU Filtration	1.4360	1.1572	0.4653
PF - SIS	2.8895	6.5271	0.5457
RMSE			
KF Unconstrained	2.3776	4.9183	0.3837
PDF Truncation	1.4826	5.4753	0.4801
SU Filtration	1.9802	1.1887	0.5694
PF - SIS	2.8453	6.4300	0.6711

Table 6.2: Comparison of Predicted Measurement Errors

The predicted measurement error of the i -th variable at time t (PME_t^i) is calculated as

$$PME_t^i = \hat{\mathbf{y}}_{t|t-1}^i - \mathbf{y}_{t|t-1}^i = \hat{\mathbf{x}}_{t|t-1}^i - \mathbf{y}_{t|t-1}^i = \hat{\mathbf{w}}_{t|t-1}^i, \quad (6.21)$$

for $t = 1, \dots, T$, superscript i goes for CPI inflation, real GDP growth and nominal interest rate and $\hat{\mathbf{x}}_{t|t-1}^i$ is a prediction of the respective state variable to the measurement one – there is a mapping of the measured variable to the relevant state variable through an identity matrix without any control variable, see equation (6.19). The PME_t^i corresponds to the prediction error term $\hat{\mathbf{w}}_{t|t-1}^i$ in the measurement equation. Based on this, the root mean square error for the

i -th measurement variable (RMSE ^{i}) is

$$\text{RMSE}^i = \sqrt{\frac{1}{T} \sum_{t=1}^T \text{PME}_t^i{}^2} = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\hat{\mathbf{x}}_{t|t-1}^i - \mathbf{y}_{t|t-1}^i \right)^2}. \quad (6.22)$$

Regarding a mean value of the prediction error, the unconstrained KF significantly outperform the rest of the methods. But it does not fulfill the non-negative real interest rate trend constraint. By imposing the required constraint, the SIS is better on average compared to the PDF truncation and the SU filtration for the measurement variables. The SU filtration is significantly worse for CPI inflation, the PDF truncation for real GDP growth. By analyzing standard deviations of the predicted errors, the situation is completely opposite - the worst performance for all three variables is connected with the SIS. Except of real GDP growth, the PDF truncation is similar to the SU model. This means that the SIS gives the smallest average predicted error, but the average magnitude of these errors is big compared to the rest methods. Using these two criteria, the SU model can be assessed as the most suitable technique.

Based on the RMSE, both the PDF truncation and the SU model evidently outperform not only the SIS, but also the unconstrained Kalman filtration. From this point, it is possible that by imposing some constraints to a system, the filtration uses additional available information, which can improve the overall results (measured by the RMSE). The RMSE of the SU filtration is comparable with the PDF truncation for CPI inflation and interest rate, but it is better for real GDP growth.

Using mean value, standard deviation and RMSE based on predicted measurement errors, the SU model seems to outperform the rest techniques.

The previous analysis is further supplemented by a visual inspection of predicted trajectories of the measurement variables, see Figure 6.12. An inappropriate method for a prediction of real GDP growth is the SIS, because the volatility of the predicted time series is significantly higher than the observed one. Contrary, the unconstrained KF and the PDF truncation capture only a trend component of the real GDP growth. Whilst all the methods except of the SU model satisfactorily follow a trend in CPI inflation and are very close to each other, the SU model is completely inappropriate. During some period the prediction of CPI inflation by the SU model does not follow the observed dynamics. In case of the nominal interest rate, all methods are close to the measured dynamics, correctly predicting a peak in 2008 and subsequent fast drop. In the last third of the sample, all methods (including the unconstrained KF) predict higher nominal interest rate to satisfy the imposed constrained. In other words, if the Czech National Bank had kept non-negative real interest rate trend (as a permanent component of the real interest rate), it would have been connected with higher nominal interest rates. This is a logical and expected result.

Before we proceed, let's check also a comparison of the prediction of the real interest rate trend in Figure 6.13. Except the unconstrained KF, the real interest rate trends are non-negative. Economically, it is intuitive that in case that there is a constraint for the real interest rate trend, the trend sticks to this constraint during periods when the constraint is active. This is a case of the SU model. However, this series is a trend and it is expected to be a smooth line. From a

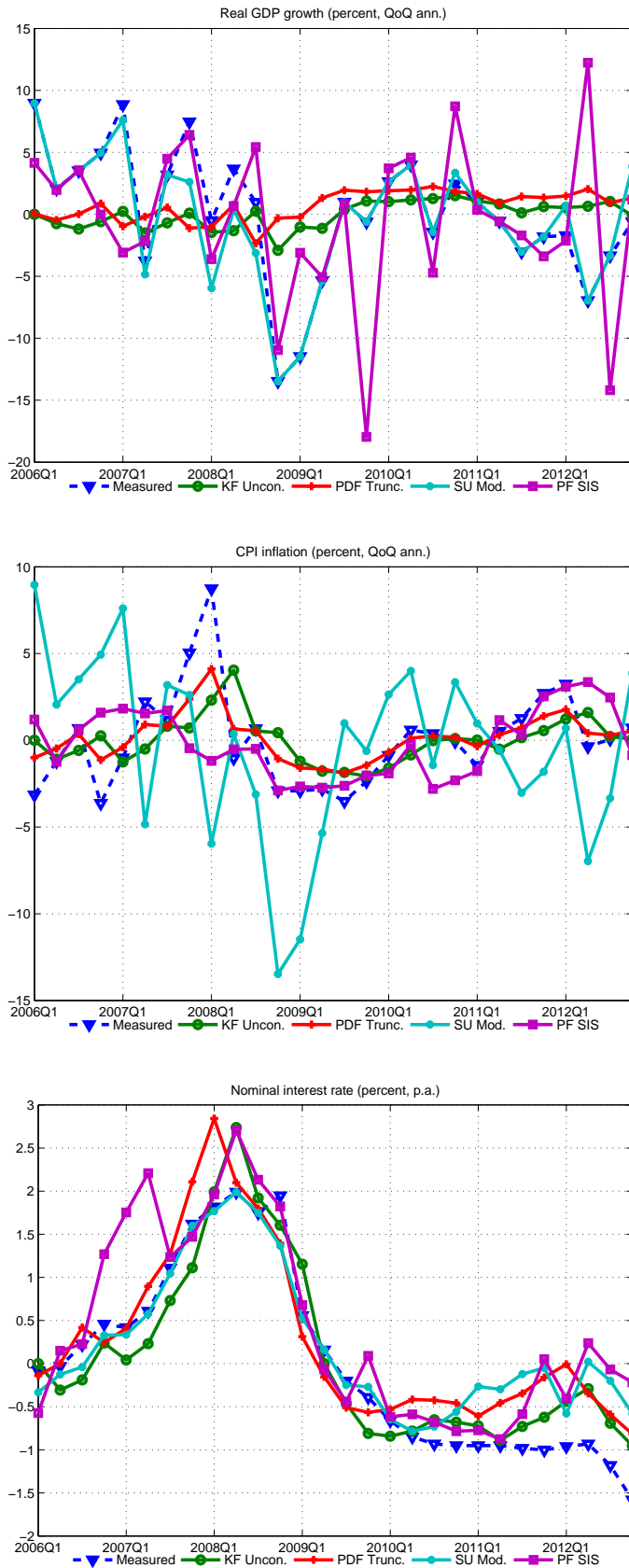


Figure 6.12: Comparison of Predicted Measurement Error (Source: author's calculation)

point of the economic theory, it is impossible to change some trend values by about 150 bp by quarter to quarter (see 2006Q3 or 2007Q3). Similar problem also faces the SIS filtration. Only the result of the PDF truncation seems to be appropriate.

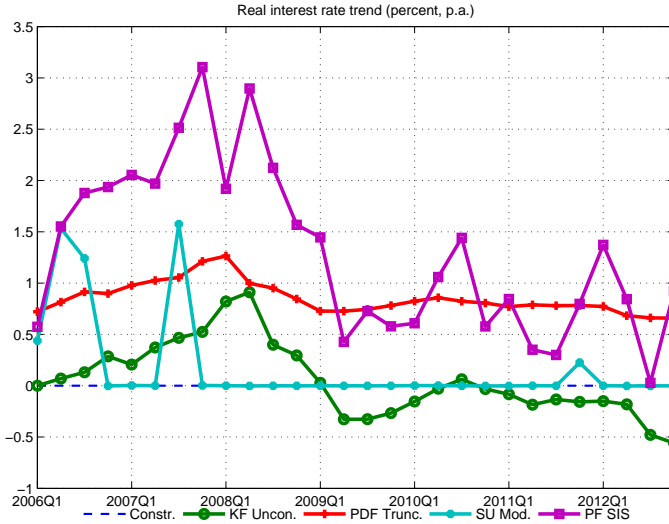


Figure 6.13: Comparison of Real Interest Rate Trend (Source: author’s calculation)

Speed of Algorithms. The PDF truncation and the SU model filtration run recursively and once the filtration at time t is done, it continues. In case of the SIS, or generally particle filters, the MC simulations are more time demanding. By imposing any constraints and resampling due to a degeneracy problem, the algorithm is incomparably slower. Additionally, due to a problem of a slow convergence, it is suitable to run extra simulations.

Based on the previously analyzed results, the PDF truncation and the SU model filtration require less computational time and the results are not substantially worse compared to the SIS outcomes.¹¹ Thus they are preferable from a point of a speed and time consumption of calculations.

Algorithm Initiation. As it was already discussed earlier, all three algorithms must be initialized by some values of particular variables.

The KF recursion, either unconstrained or constrained, starts with $\hat{\mathbf{x}}_{1|0}$, which denotes a forecast of \mathbf{x}_1 based on no observations with the associated MSE expressed by $\mathbf{P}_{1|0}$. Generally, if the initial state is not available, it is replaced with an arbitrary option of the analyst’s best guess to the initial value and a guess about its confidence, see Subsection 2.1.1. Not to make any algorithm advantageous, the initial state is set to zero for all its elements (a mean value for all the states, because the model is specified in a gap form) and the uncertainty is based on the estimation of the covariance matrix $\mathbf{P}_{1|0}$ from the KF smoother introduced in Subsection 6.1.3. Additionally, it is necessary to calibrate standard deviations of structural shocks and measurement errors. This is done relatively to each other (not in absolute terms), but it requires an iterative process based

¹¹On top of that, by a 2000-run MC simulation is not still ensured a convergence of the chain, however, this takes at about one hour. See Subsection 6.2.3.

on model structure, stylized facts about the modeled economy, checking filtered results and judgments as well.

The same initialization is required for the SIS method - an initial value for the state vector (again all elements of the vector are set to zero) and calibration of the standard deviations of structural shocks. There is no setting of an initial confidence about the state vector. It is expressed by a calibration of the standard deviations of the structural shocks and measurement errors. This is a big advantage, but a value of the threshold in case of the degeneracy problem and a number of MC simulations for convergence must be specified instead.

The SU model is estimated using the technique described in Section 4.2. To successfully run a non-linear programming algorithm, a starting point for \mathbf{X} of the optimization has to be set appropriately. Improper setting causes numerical instability. However, in case of the linear algorithm, nothing else is required to be set for an initialization of the algorithm.

6.4 Discussion of Results

Based on the results in this chapter, the SU model filtration technically outperforms the PDF truncation and the SIS methods. Results of the SU model are comparable to the KF with the PDF truncation and to the SIS filtration. They are even better in case of the RMSE for the predicted measurement errors. An advantage of the SU model is that no initialization of the algorithm was required and just a very rough knowledge about the actual value of noise boundaries is required only.

On the other hand, economic background and interpretation behind the SU model filtration is weak and beat by the SIS and especially by the PDF truncation. The model is designed to be a monetary policy model of the Czech economy, focusing especially on CPI and nominal interest rate dynamics. In case of CPI inflation, the unsatisfactory results of the SU model occur mostly when an outage is placed on the sharp turns of the trajectory - e.g. a peak in inflation during 2008 and 2011. Reflecting these outcomes, economic analysts would very likely prefer the PDF truncation as an appropriate method for the filtration.

To conclude the comparison of the methods, it is necessary to stress that the results can be influenced by several aspects. Only three statistical criteria were introduced - mean, standard deviation and RMSE of the predicted error for the measurement variables. They do not offer unambiguous outcomes and introducing different criteria might change preferences about a suitable method. There is no deep analysis of an effect of initial conditions - adjusting the initial conditions to every particular method can improve overall results. A poor initial guess can be typical in many practical situations, however, it can cause a slow initial convergence and poor accuracy especially in case of MC simulations. It is convenient to check a robustness of results to a different set of initial conditions. Due to a simplification, only one set of the initial conditions were used.

Finally, all three algorithms use different techniques for filtration and estimation of an unobserved state vector. This may require a slightly different calibration or specification of the semi-structural model to offer appropriate results. For the comparison, the structure of the model, calibration of parameters and other settings is exactly the same one. It would be useful to discuss some properties of

the calibration and designed model structure in more details and point out some possible drawbacks and potential areas for an improvement.

Conclusion

The thesis introduces a set of methods to identify unobserved variables using a constrained dynamic system. Three of them are described in details and subsequently applied to economic data and semi-structural simple macroeconomic model for monetary policy analysis. The thesis does not present a novel approach to the filtration techniques, but rather offers a general overview of algorithms with a relatively wide literature survey including up-to-date techniques.

An application part deals with only a linear model and a linear constraint although several methods for non-linear problems are described. Introducing non-linear filtration and/or non-linear constraints is very often based on linearization and thus completes an overview of the techniques. The used codes in the application part can be easily extended to the non-linear case.

The thesis considers a problem of a state-space representation of a model and an application of a Kalman filter-type estimator to a system with and without state constraints. This is extended by imposing a non-linear constraints and/or non-linear KF techniques. The thesis proposes several different ways of constraints. A detailed derivation of the KF is motivated by a fact that the KF is one of the most widely used methods for estimation due to its simplicity, optimality, tractability and robustness. However, the application of the KF to non-linear systems and/or constraints is challenging. Moreover, this area has not been satisfactorily explored so far.

In many situations, not only the economic ones, certain information about a dynamical system is most naturally represented using constraints. Including such information can compensate for model approximation errors and can improve the state estimates. Strategies for including equality constraints can be found in the literature. However, handling inequality constraints is a much harder problem due to the complexity of identifying those constraints that are active at the solution. Nonetheless, estimates using inequality constraints may significantly change results, as it was discussed in the application part of the thesis.

The second proposed algorithm, the SU model with boundary noise, is an alternative to the KF based algorithms. It is simple to perform and it needs no demanding initial setting. This feature can be important for practitioners. With proposed model, they avoid a complicated theoretical setting that is necessary for a successful running of KF based algorithms. The thesis applied this algorithm to an economic problem and filtration of macroeconomic data for the first time. Based on the application part, the algorithm exhibits two glamorous characteristics. First, it allows an estimation of the innovation range and second, it allows (without excessive computational demands) to respect prior bounds on model parameters and states.

The third introduced method is a general approach for estimation of nonlinear dynamic process systems. This approach is based on a rigorous Bayesian formulation of estimation problems that uses sequential Monte Carlo sampling to propagate all information recursively, while minimizing assumptions about the system. The resulting approach does not rely on common assumptions of Gaussian or fixed-shape distributions, which can be easily violated in (non-) linear dynamic systems. Therefore, this approach is expected to handle broader range

of problems and result in more accurate estimation than existing approaches. In practice it means, that a comparison with the SU model filtration can be more rigorous using an assumption of uniformly distributed structural disturbances. This is one of a proposed direction for the future work.

The application part compares the unconstrained KF, the KF with the PDF truncation, the SU model filtration and the SIS with resampling and an acceptance/rejection algorithm. The three later techniques enforce the inequality constraint for one of state variables. A comparison with respect to several aspects is done. Results confirm the fact that an appropriate method depends on an aim of the filtration. From a technical point, the SU model seems to be the most proper one. Contrary, the PDF truncation is the easiest one for an economic interpretation and the closest one to an economic intuition. Because of a relatively simple model framework with a Gaussian structural shocks and measurement errors, the SIS has not shown its full capability and power to overweight the computational burden. This confirms a fact that a proper method reflects an aim of the filtration, depends on a measurement variable trajectory shape and measurement data and a final goal of an analysis. None of the used methods outperform the rest of them in majority of the analyzed aspects.

There are a number of directions for future work which formed limits for the previous comparison.

First, a prediction step of the filtration was used although it is more common in practice to use either an update or a smoothed step of the filtration. This very likely results in a different dynamics of the state variables and measurement errors identified by the dynamic system.

Second, with an exception of the SU model, the prediction of the real interest rate trend, as a constrained state variable, resulted in a non-zero value filtration, which is out of the economic intuition. It would be interesting to check an outcome of imposing either zero equality constraint or upper inequality constraints in periods when the unconstrained filtration for the real interest rate trend was negative. More advanced exercise can be connected to an introduction of the non-linear inequality constraints, e.g. a penalty function in case of a negative real interest rate trend.

Third and closely connected to the previous area, it would be useful to compare the performance of different algorithms to see if an over-constrained representation provides benefits or not.

Fourth, the used model was a monetary policy model, but it lacks of the exchange rate dynamics. It means that the model should be extended into a form of an open economy model. The nominal exchange rate can be another channel (under a direct control of the authority) through which the monetary authority can influence the domestic economy to avoid a long-term negative interest rate. It means that the non-negative real interest rate need not be necessarily fully offset by an increase in the nominal interest rate and a fiscal expansion through the domestic aggregate demand increase.

Fifth, it has already been stressed that the imposed inequality constrain can be time-varying to better fit a changing structure of the economy. This reflects mainly a change of the inflation target which was reduced from 3 to 2 percent in 2010Q1. A code adjustment should be relatively simple in this situation.

Finally, a given structure and calibration of the employed model raises ques-

tion as to what is the best structural representation and calibration of the model and what is an impact of this on estimator performance.

To conclude, the thesis can offer a fundamental overview of the filtering techniques using inequality constraints and it can also contribute to the recent discussion about constraints faced by the monetary policy. Using inequality constraints or non-linear systems with non-linear constraints prove to be useful, because economic behavior of developed countries during and after the global financial crisis exhibits significant non-linear features, e.g. non-negative nominal interest rate, non-linearly changing country and governmental risk premium with respect to the public debt development, or coordinated monetary and fiscal policies expansion inefficiently influencing the real economic performance and unemployment.

Bibliography

- [1] AGATE, Craig S. and SULLIVAN, Kevin J. *Road-Constrained Target Tracking and Identification Using a Particle Filter*. Working Paper. Toyon Research Corporation, California, 2003.
- [2] ALOUANI, Ali T. and BLAIR, William D. *Use of a Kinematic Constraint in Tracking Constant Speed, Maneuvering Targets*. IEEE Transactions on Automatic Control, 1993. Vol. 38, Iss. 7.
- [3] ANDĚL, Jiří. *Basis of the Mathematical Statistics*. 2nd edition. In Czech, original title *Základy matematické statistiky*. Prague: Matfyzpress, 2007. ISBN 80-7378-001-1.
- [4] ANDERSON, Brian D. O. and MOORE, John B. *Optimal Filtering*. Republic of Dover edition. New York: Dover Publications, 2005. ISBN 0-486-43938-0.
- [5] ANDRLE, Michal, HLEDÍK, Tibor, KAMENÍK, Ondra and VLČEK, Jan. *Implementing the New Structural Model of the Czech National Bank*. Working Paper 2009/2. Czech National Bank, Research Department, 2009.
- [6] ARULAMPALAM, Sanjeev M., MASKELL, Simon and CLAPPO, Tim. *A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking*. IEEE Transactions on Signal Processing, 2002. Vol. 50, No. 2.
- [7] BELL, Bradley M., BURKE, James V. and GILLONETTO, Gianluigi. *An Inequality Constrained Nonlinear Kalman-Bucy Smoother by Interior Point Likelihood Maximization*. Automatica, 2009. Vol. 45, Iss. 1.
- [8] BERG, Andrew, KARAM, Philippe and LAXTON, Douglas. *Practical Model-Based Monetary Policy Analysis: A How-To Guide*. IMF Working Paper 06/81. International Monetary Fund, 2006.
- [9] BERGER, James, O. *Statistical Decision Theory and Bayesian Analysis*. 2nd edition with 23 illustrations. New York, US: Springer, 1985. ISBN 0-387-96098-2.
- [10] BOERS, Yvo and DRIESSEN, Hans. *Particle Filter Track-before-Detect Application Using Inequality Constraints*. IEEE Transactions on Aerospace and Electronic Systems, 2005. Vol. 41, Iss. 4.
- [11] BOYD, Stephen and VANDENBERGHE, Lieven. *Convex Optimization*. 7th printing with corrections. Cambridge, US: Cambridge University Press, 2009. ISBN 978-0-521-83378-3.
- [12] BROWN, Robert G. and HWANG, Patrick Y. C. *Introduction to Random Signals and Applied Kalman Filtering (with MATLAB Exercises and Solutions)*. 3rd edition. Wiley, US: John Wiley & Sons, 1997. ISBN 0-471-12839-2.
- [13] CAINES, Peter, E. *Linear Stochastic Systems*. 99th edition. Wiley, US: John Wiley & Sons, 1988. ISBN 0-471-08101-9.

- [14] CHEN, Wen-shiang. *Bayesian Estimation by Sequential Monte Carlo Sampling for Nonlinear Dynamic Systems*. Dissertation, The Ohio State University, 2004.
- [15] CHEN, Wen-shiang, BAKSHI, Bhavik R., GOEL, Prem K. and CHIZECK, Sridhar. *Bayesian Estimation via Sequential Monte Carlo Sampling - Unconstrained Nonlinear Dynamic Systems*. Industrial & Engineering Chemistry Research, 2004. Vol. 43, Iss. 14.
- [16] CHEN, Zhe. *Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond*. Manuscript. McMaster University, 2003.
- [17] CHIA, Tien L., CHOW, Peter and CHIZECK, Howard J. *Recursive Parameter Identification of Constrained Systems: an Application to Electrically Stimulated Muscle*. IEEE Transactions on Biomedical Engineering, 1991. Vol. 38, Iss. 5.
- [18] CIPRA, Tomáš. *Financial Econometrics*. 1st edition. In Czech, original title *Finanční ekonometrie*. Prague: Ekopress, 2008. ISBN 978-80-86929-43-9.
- [19] CRISAN, Dan and DOUCET, Arnaud. *A Survey of Convergence Results on Particle Filtering Methods for Practitioners*. IEEE Transactions on Signal Processing, 2002. Vol. 50, No. 3.
- [20] CZECH NATIONAL BANK. *The Czech National Bank's Forecasting and Policy Analysis System*. Working Paper. Edited by Coats, W., Laxton, D. and Rose, D. Czech National Bank, 2003.
- [21] DORAN, Howard. *Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions*. The review of Economics and Statistics, 1992. Vol. 74, Iss. 3.
- [22] DURRANT-WHYTE, Hugh. *Introduction to Estimation and the Kalman Filtering*. Textbook, version 2.2. Australian Centre for Field Robotics, The University of Sydney, 2001.
- [23] DE GEETER, Jan, VAN BRUSSEL, Hendrik, DE SCHUTTER, Joris and DECRETON, Marc. *A Smoothly Constrained Kalman Filter*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997. Vol. 19, No. 10.
- [24] JACOBS, Oliver, L. R. *Practical Methods of Optimization*. 2nd edition. New York, US: Oxford University Press, 1993. ISBN 0-19-856249-7.
- [25] DE JONG, Piet. *The Likelihood for a State Space Model*. Biometrika, 1988. Vol. 75, No. 1.
- [26] DE JONG, Piet. *Smoothing and Interpolation with the State-Space Model*. Journal of the American Statistical Association, 1989. Vol. 84, No. 408.
- [27] FLETCHER, Roger. *Practical Methods of Optimization*. 2nd edition. Wiley, US: John Wiley & Sons, 2000. ISBN 978-0-471-49463-8.

- [28] FRANCKE, Marc K., KOOPMAN, Siem J. and DE VOS, Aart F. *Likelihood Functions for State Space Models with Diffuse Initial Conditions*. Tinbergen Institute Discussion Paper TI 2008-040/4. Tinbergen Institute, 2008.
- [29] GOLDFELD, Stephen M. and QUANDT, Richard E. *A Markov Model for Switching Regressions*. Journal of Econometrics, 1973. Vol. 1, Iss. 1.
- [30] GORDON, Neil J., SALMOND, David J. and SMITH, Adrian F. M. *Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation*. IEE Proceeding - F, 1993. Vol. 140, No. 2.
- [31] GUPTA, Nachi and HAUSER, Raphael. *Kalman Filtering with Equality and Inequality State Constraints*. Report no. 07/18. Oxford University Computing Laboratory, Numerical Analysis Group, 2007.
- [32] HAMILTON, James C. *Time Series Analysis*. 1st edition. Princeton, New Jersey: Princeton University Press, 1994. ISBN 0-691-04289-6.
- [33] HAMILTON, James C. *State-Space Models*. Handbook of Econometrics, 1994. Vol. IV.
- [34] HARVEY, Andrew C. *Forecasting, Structural Time Series Models and the Kalman Filter*. Reprinted. Cambridge, England: Cambridge University Press, 1990. ISBN 0-521-32196-4.
- [35] HAVLENA, Vladimír and ŠTECHA, Jan. *Modern Control Theory*. Textbook, Czech Technical University in Prague, Faculty of Electrical Engineering. In Czech, original title *Moderní teorie řízení*. Prague: Ediční středisko ČVUT, 1999.
- [36] JULIER, Simon J. and LAVIOLA, Joseph J. (Jr.) *On Kalman Filtering with Nonlinear Equality Constraints*. IEEE Transactions on Signal Processing, 2007. Vol. 55, No. 6.
- [37] JULIER, Simon J. and UHLMANN, Jeffrey K. *Unscented Filtering and Nonlinear Estimation*. Proceedings of the IEEE, 2004. Vol. 92, No. 3.
- [38] JULIER, Simon J., UHLMANN, Jeffrey K. and DURRANT-WHYTE, Hugh, F. *A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators*. IEEE Transactions on Automatic Control, 2000. Vol. 45, No. 3.
- [39] JULIER, Simon J. and UHLMANN, Jeffrey K. *A General Method of Approximating Nonlinear Transformations of Probability Distribution*. TechReport. Department of Engineering Science, University of Oxford, 1996.
- [40] KALMAN, Rudolf E. *A New Approach to Linear Filtering and Prediction Problems*. AMSE - Journal of Basic Engineering, 1960. Vol. 82, Series D.
- [41] KANAZAWA, Keiji, KOLLER, Daphne and RUSSELL, Stuart J. *Stochastic Simulation Algorithms for Dynamic Probabilistic Networks*. Proceedings of the Eleventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-95), 1995.

- [42] KÁRNÝ, Miroslav and PAVELKOVÁ, Lenka. *Projection-based Bayesian Recursive Estimation of ARX Model with Uniform Innovations*. Systems and Control Letters, 2007. Vol. 56, Iss. 9-10.
- [43] KÁRNÝ, Miroslav, BOHM, Josef, GUY, Tatiana V., JIRSA, Ladislav, NAGY, Ivan, NEDOMA, Petr and TESAŘ, Ludvík. *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. 2006 Edition. London, England: Springer - Verlag, 2005. ISBN 1-85233-928-4.
- [44] KO, Sangho and BITMEAD, Robert R. *State Estimation for Linear Systems with State Equality Constraints*. Automatica, 2007. Vol. 43, Iss. 8.
- [45] KOHN, Robert and ANSLEY, Craig F. *Fixed Interval Estimation in State Space Models when Some of the Data Are Missing or Aggregated*. Biometrika, 1983. Vol. 70, No. 3.
- [46] KOHN, Robert and ANSLEY, Craig F. *Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data*. Journal of American Statistical Association, 1986. Vol. 81, No. 395.
- [47] KONG, Augustine, LIU, Jun S. and WONG, Wing H. *Sequential Imputations and Bayesian Missing Data Problems*. Journal of American Statistical Association, 1994. Vol. 89, No. 425.
- [48] KUNCH, Hans R. *Recursive Monte Carlo Filter: Algorithms and Theoretical Analysis*. Annals of Statistics, 2005. Vol. 35, No. 5.
- [49] KYRIAKIDES, Ioannis, MORRELL, Darryl and PAPANDREOU-SUPPAPPOLA, Antonia. *A Particle Filtering Approach to Constrained Motion Estimation in Tracking Multiple Targets*. Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers in Pacific Grove, California, 28 October – 1 November, 2005.
- [50] KYRIAKIDES, Ioannis, MORRELL, Darryl and PAPANDREOU-SUPPAPPOLA, Antonia. *Multiple Target Tracking with Constrained Motion Using Particle Filtering Methods*. 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 13 – 15 December, 2005.
- [51] LANG, Lixin, CHEN, Wen-shiang, BAKSHI, Bhavik R., GOEL, Prem K. and UNGRALA, Shidhar. *Bayesian Estimation via Sequential Monte Carlo Sampling – Constrained Dynamic Systems*. Automatica, 2007. Vol. 43, Iss. 9.
- [52] LEFEBVRE, Tine, BRUYNINCKX, Herman and DE SCHUTTER, Joris. *Comment on “A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators”*. IEEE Transactions on Automatic Control, 2002. Vol. 47, Iss. 8.
- [53] LUENBERGER, David, G. and YE, Yinyu. *Linear and Nonlinear Programming*. 3rd edition. New York, U.S.: Springer, 2008. ISBN 978-0-387-74502-2.
- [54] MACCORMICK, John and BLAKE, Andrew. *A Probabilistic Exclusion Principle for Tracking Multiple Objects*. International Journal of Computer Vision, 2000. Vol. 39, No. 1.

- [55] MAHATA, Kaushik and SODERSTROM, Torsten. *Improved Estimation Performance Using Known Linear Constraints*. Automatica, 2004. Vol. 40, Iss. 8.
- [56] MASSICOTTE, Daniel, MORAWSKI, Roman Z., and BARWICZ, Andrzej. *Incorporation of a Positivity Constraint into a Kalman-Filter-Based Algorithm for Correction of Spectrometric Data*. IEEE Transactions on Instrumentation and Measurement, 1995. Vol. 44, Iss. 1.
- [57] MAYBECK, Peter S. *Stochastic Models, Estimation, and Control*. Volume 1. New York, U.S.: Academic Press, 1979. ISBN 0-12-480701-1.
- [58] MEINHOLD, Richard J. and SINGPURWALLA, Nozer D. *Understanding the Kalman Filter*. The American Statistician, 1983. Vol. 37, No. 2.
- [59] DEL MORAL, Pierre. *Nonlinear Filtering: Interacting Particle Resolution*. Markov Processes and Related Fields, 1996. Vol. 2, No. 4.
- [60] PAVELKOVÁ, Lenka. *Nonlinear Bayesian State Filtering with Missing Measurement and Bounded Noise and its Application to Vehicle Position Estimation*. Kybernetika, 2011. No. 3, Vol. 47.
- [61] PAVELKOVÁ, Lenka. *Vehicle Position Estimation Using GPS/CAN Data Based on Nonlinear Programming*. Proceedings of the IASTED International Conference, Cambridge, United Kingdom, 11 - 13 July, 2011.
- [62] PAVELKOVÁ, Lenka. *State Estimation with Missing Data and Bounded Uncertainty*. Research Report No. 2296. Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 2011.
- [63] PAVELKOVÁ, Lenka. *Approximate On-line Estimation of Uniform State Model with Application on Traffic Data*. Regular Paper. International Workshop on Assessment and Future Directions of NMPC, Pavia, Italy, 5 - 9 September, 2008.
- [64] PAVELKOVÁ, Lenka. *Estimation of Models with Uniform Innovations and its Application on Traffic Data*. PhD Thesis. Faculty of Transportation Sciences, Czech Technical University in Prague, 2008.
- [65] RAO, Christopher V. and RAWLINGS, James B. *Constrained Process Monitoring: Moving-Horizon Approach*. AIChE Journal, 2005. Vol. 48, Iss. 1.
- [66] RAO, Christopher V., RAWLINGS, James B. and LEE, Jay H. *Constrained Linear State Estimation - A Moving Horizon Approach*. Automatica, 2001. Vol. 37, Iss. 10.
- [67] RIBERIO, Isabel M. *Kalman and Extended Kalman Filters: Concept, Derivation and Properties*. Working Paper. Instituto Superior Técnico, Institute for Systems and Robotics, 2004.
- [68] RHODES, Ian B. *A Tutorial Introduction to Estimation and Filtering*. IEEE Transactions on Automatic Control, 1971. Vol. AC-16, No. 6.

- [69] ROTHENBERG, Thomas, J. *Identification in Parametric Model*. *Econometrica*, 1971. Vol. 39, No. 3.
- [70] LE ROUX, Joel. *An Introduction to Kalman Filtering: Probabilistic and Deterministic Approaches*. Working Paper. University of Nice, 2003.
- [71] SAYED, Ali, H. *A Framework for State-Space Estimation with Uncertain Models*. *IEEE Transactions on Automatic Control*, 2001. Vol. 46, No. 7.
- [72] SIMON, Dan. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. 1st edition. Wiley, US: John Wiley & Sons, 2006. ISBN 0-471-70858-5.
- [73] SIMON, Dan. *Kalman Filtering with State Constraints: A Survey of Linear and Nonlinear Algorithms*. *IET Control Theory and Applications*, 2010. Vol. 4, Iss. 8.
- [74] SIMON, Dan and SIMON, Donald L. *Kalman Filtering with Equality Constraints for Turbofan Engine Health Estimation*. *IEEE Proceeding – Control Theory and Applications*, 2006. Vol. 153, No. 3.
- [75] SIMON, Dan and CHIA, Tien L. *Kalman Filtering with State Equality Constraints*. *IEEE Transactions on Aerospace and Electronic Systems*, 2002. Vol. 38, No. 1.
- [76] SIRCOULOMB, Vincent, HOBLOS, Ghaleb, CHAFOUK, Houcine and RAGOT, José. *State Estimation under Nonlinear State Equity Constraints. A Tracking Problem*. 16th Mediterranean Conference on Control and Automation - MED in Ajaccio, France, 25 – 27 June, 2008.
- [77] SPONG, Mark W., HUTCHINSON, Seth and VIDYASAGAR, M. *Robot Modeling and Control*. 1st edition. Wiley, US: John Wiley & Sons, 2005. ISBN 0-471-64990-2.
- [78] ŠTECHA, Jan and HAVLENA, Vladimír. *Dynamic Systems Theory*. Textbook, Czech Technical University in Prague, Faculty of Electrical Engineering. In Czech, original title *Teorie dynamických systémů*. Prague: Ediční středisko ČVUT, 1995.
- [79] STENGEL, Robert, F. *Optimal Control and Estimation*. 1st edition. New York, US: Dover Publications, 1994. ISBN 0-486-68200-5.
- [80] VAN SUNTUM, Ulrich, KAPTAN, Metin and ILGMANN, Cordelius. *Reducing the Lower Bound on Market Interest Rates*. *Economic Analysis and Policy*, 2011. Vol. 41, No. 2.
- [81] TEIXEIRA, Bruno O. S., CHANDRASEKAR, Jaganath, TORRES, Leonardo A. B., AGUIRRE, Luis A. and BERNSTEIN, Dennis S. *State Estimation for Linear and Non-Linear Equality-Constrained Systems*. *International Journal of Control*, 2009. Vol. 82, No. 5.

- [82] TEIXEIRA, Bruno O. S., TORRES, Leonardo A. B., AGUIRRE, Luis A. and BERNSTEIN, Dennis S. *Unscented Filtering for Interval-Constrained Nonlinear Systems*. Proceedings of the 47th IEEE Conference on Decision and Control in Cancun, Mexico, December 9 – 11, 2008.
- [83] TEREJANU, Gabriel A. *Extended Kalman Filter Tutorial*. Tutorial. Department of Computer Science and Engineering, University at Buffalo, 2008.
- [84] VACHHANI, Pramod, NARASIMHAN, Shankar and RENGASWAMY, Raghunathan. *Robust and Reliable Estimation via Unscented Recursive Nonlinear Dynamic Data Reconciliation*. Journal of Process Control, 2006. Vol. 16, Iss. 10.
- [85] VACHHANI, Pramod, RENGASWAMY, Raghunathan, GANGWAL, Vikrant and NARASIMHAN, Shankar. *Recursive Estimation in Constrained Nonlinear Dynamical Systems*. AIChE Journal, 2005. Vol. 51, Iss. 3.
- [86] WANG, L.-S., CHIANG, Y.-T. and CHANG, F.-R. *Filtering Method for Nonlinear Systems with Constraints*. IEE Proceedings - Control Theory and Applications, 2002. Vol. 149, No. 6.
- [87] WATSON, Mark W. *Recursive Solution Methods for Dynamic Linear Rational Expectations Models*. Journal of Econometrics, 1989. Vol. 41, Iss. 1.
- [88] WELCH, Greg and BISHOP, Gary. *An Introduction to the Kalman Filter*. Working Paper TR 95-041. University of North Carolina at Chapel Hill, Department of Computer Science, 2006.
- [89] WELLING, Max. *The Kalman Filter*. Classnotes in Machine Learning. California Institute of Technology, 2012.
- [90] YANG, Chun and BLASCH, Erik. *Kalman Filtering with Nonlinear State Constraints*. IEEE Transactions on Aerospace and Electronic Systems, 2008. Vol. 45, Iss. 1.

Appendix

This part contains technical overview and a core of codes used in the application part of the thesis.

Appendix 1: Model Code in Iris

A complete model code including declaration of all variables, parameters, and shocks has the following form.

```
!variables:transition
'Real GDP (100*log)' l_y
'Potential output (100*log)' l_y_tnd
'Output gap (percent)' l_y_gap
'Real GDP growth (percent, QoQ ann.)' dl_y
'Potential output growth (percent, QoQ ann.)' dl_y_tnd
'CPI inflation (percent, QoQ ann.)' dl_cpi
'Nominal interest rate (percent, p.a.)' i
'Real interest rate (percent, p.a.)' r
'Real interest rate trend (percent, p.a.)' r_tnd

!variables:measurement
'CPI inflation (percent, QoQ ann.)' obs_dl_cpi
'Real GDP growth (percent, QoQ ann.)' obs_dl_y
'Nominal interest rate (percent, p.a.)' obs_i

!shocks:transition
shock_l_y_gap
shock_dl_cpi
shock_i
shock_dl_y_tnd
shock_r_tnd

!shocks:measurement
meas_dl_cpi
meas_dl_y
meas_i

!parameters
a1 = 0.6;
a2 = -0.3;
b1 = 0.5;
b2 = 0.3;
c1 = 0.8;
c2 = 0.2;
c3 = 0.1;
d1_dl_y_tnd = 0.7;
```

```

e1_r_tnd = 0.8;

std_shock_l_y_gap = 1.4;
std_shock_dl_cpi = 1.5;
std_shock_i = 0.4;
std_shock_dl_y_tnd = 1.0;
std_shock_r_tnd = 0.7;
std_meas_dl_cpi = 0.75;
std_meas_dl_y = 0.75;
std_meas_i = 0.25;

!equations:transition
l_y_gap = a1*l_y_gap{-1} + a2*(r{-1} - r_tnd{-1}) + shock_l_y_gap;
dl_cpi = b1*dl_cpi{-1} + b2*l_y_gap + shock_dl_cpi;
i = c1*i{-1} + (1-c1)*((r_tnd + dl_cpi) + c2*dl_cpi + c3*l_y_gap)
    + shock_i;
r = i - dl_cpi;
l_y = l_y_tnd + l_y_gap;
dl_y = (l_y - l_y{-1})*4;
dl_y_tnd = (l_y_tnd - l_y_tnd{-1})*4;
dl_y_tnd = d1*dl_y_tnd{-1} + shock_dl_y_tnd;
r_tnd = e1_r*r_tnd{-1} + shock_r_tnd;

!equations:measurement
obs_dl_cpi = dl_cpi + meas_dl_cpi;
obs_dl_y = dl_y + meas_dl_y;
obs_i = i + meas_i;

```

Appendix 2: Matlab Code for PDF Truncation

The core of a calculation of the PDF truncation is the following. It requires the *MGS* function to calculate the modified Gram-Schmidt transformation. The code deals only lower bound inequalities.

```

for t = 1:length(hdata(1,:))
    z = hdata(:,t);
    % Run the Kalman filter.
    P0 = A * P0 * A' + Q;
    x = A * x;
    K = P0 * H' * inv(H * P0 * H' + R);
    x = x + K * (z - H * x);
    P0 = (eye(n, n) - K * H) * P0;
    % Constrained filtering via PDF truncation
    P_trunc = P0;
    x_trunc = x;
    for k = 1 : r
        [ U_trunc, W_trunc, V_trunc ] = svd(P_trunc);
    end
end

```

```

T_trunc = U_trunc;
TT = T_trunc * T_trunc';
% Compute the mod. Gram-Schmidt transf. S*A_mgs=[W_mgs;0]:
% A_mgs is a given n x m matrix,
% S is an orthogonal n x n matrix, and
% W_mgs is an m x m matrix.
A_mgs = sqrt(W_trunc) * T_trunc' * D(k,:)' ;
[ W_mgs, S ] = MGS(A_mgs);
S = S * sqrt(D(k,:) * P_trunc * D(k,:))' / W_mgs;
c_trunc = (d(k) - D(k,:) * x_trunc) / sqrt(D(k,:) * ...
    P_trunc * D(k,:))';
d_trunc = (Inf - D(k,:) * x_trunc) / sqrt(D(k,:) * ...
    P_trunc * D(k,:))';
alpha = sqrt(2/pi) / (erf(d_trunc/sqrt(2)) - ...
    erf(c_trunc/sqrt(2)));
mu = alpha * (exp(-c_trunc^ 2/2) - exp(-d_trunc^ 2/2));
sigma2 = alpha * (exp(-c_trunc^ 2/2) * (c_trunc - ...
    2 * mu) - exp(-d_trunc^ 2/2) * (d_trunc - 2 * mu)) + ...
    mu^ 2 + 1;
z_trunc = zeros(size(x_trunc));
z_trunc(1) = mu;
Z_cov = eye(length(z_trunc));
Z_cov(1,1) = sigma2;
x_trunc = T_trunc * sqrt(W_trunc) * S' * z_trunc + x_trunc;
P_trunc = T_trunc * sqrt(W_trunc) * S' * Z_cov * S * ...
    sqrt(W_trunc) * T_trunc';
end
% Compute prediction step
x_pred = A * x;
x_trunc_pred = A * x_trunc;
% Save data in arrays
x_array = [x_array x];
x_trunc_array = [x_trunc_array x_trunc];
x_pred_array = [x_pred_array x_pred];
x_trunc_pred_array = [x_trunc_pred_array x_trunc_pred];
end

```

Appendix 3: Matlab Code for SU Model Filtration

The core of a code for the SU model is following. It requires to link the Optimization Toolbox in Matlab and the *lsuestlin* function for the linear SU model estimation.

```

DATA(ychns,TIME)=hdata(:,1);
% SU model estimation
[LsuMem,xest,rxest,ryest,Eest,Fest,lagrange] = lsuestlin(LsuMem);

```

```

Lagrange{TIME}=lagrange;
lenx=size(xest,2);
jj=0;
StatesE(:,TIME)=LsuMem{1}.state;
for ii=TIME:-1:max(TIME-lenx+1,1)
    StatesEcell{ii}=[StatesEcell{ii} xest(:,lenx-jj)];
    jj=jj+1;
end
% Estimated noise boundaries
xrE(:,TIME)=rxest;
yrE(:,TIME)=ryest;
% Estimated parameters if necessary
EE{TIME}=Eest;
FE{TIME}=Fest;
% Update of the predictor
Pred{1}.Ec=Eest;
Pred{1}.Fc=Fest;
Pred{1}.xr=rxest;
Pred{1}.yr=ryest;
% Output prediction with estimated states
[Pred,y,x]=lsusimul(Pred,DATA(uchns,TIME),LsuMem{1}.state);
DATA(ychns_pred,TIME)=y;
% On-line state and parameters and noise bounds estimation
for TIME = 3:ndat;
    mem1=min(mem,TIME-1);
    DATA(ychns,TIME)=hdata(:,TIME-1);
    [LsuMem,xest,rxest,ryest,Eest,Fest,lagrange] = lsuestlin(LsuMem);
    Lagrange{TIME}=lagrange;
    lenx=size(xest,2);
    jj=0;
    % Estimated newest state
    StatesE(:,TIME)=LsuMem{1}.state;
    for ii=TIME:-1:max(TIME-lenx+1,1)
        % All estimated states
        StatesEcell{ii}=[StatesEcell{ii} xest(:,lenx-jj)];
        jj=jj+1;
    end
    % Estimated noise boundaries
    xrE(:,TIME)=rxest;
    yrE(:,TIME)=ryest;
    % Estimated parameters
    EE{TIME}=Eest;
    FE{TIME}=Fest;
    % Update of the predictor
    Pred{1}.Ec=Eest;
    Pred{1}.Fc=Fest;
    Pred{1}.xr=rxest;
    Pred{1}.yr=ryest;

```

```

% Output prediction with states estimated
[Pred,y,x]=lsusimul(Pred,DATA(uchns,TIME),LsuMem{1}.state);
DATA(ychns_pred,TIME)=y;
% Computation of state estimates characteristic
StatesEmean(:,TIME-mem1)= mean(StatesEcell{TIME-mem1},2);
StatesEstd(:,TIME-mem1) = std(StatesEcell{TIME-mem1},0,2);
xxx=StatesEcell{TIME-mem1};
for i=1:stsize
    % Sort state estimates in ascending order
    xxxs=sort(xxx(i,:),2,'ascend');
    nn=length(xxxs);
    % The smallest estimate
    StatesElowint(i,TIME-mem1)=xxxs(1,1);
    smin=xxxs(1,1);
    % The biggest estimate
    StatesEupint(i,TIME-mem1) =xxxs(1,length(xxxs));
    smax=xxxs(1,length(xxxs));
    % Computation of rho expectation
    Ero=comput_ero(smin,smax,nn);
    Eroall(i,TIME-mem1)=Ero;
    % Computation of x expectation
    Ex = comput_ex(smin,smax,nn);
    Exall(i,TIME-mem1)=Ex;
    % Lower boundary of interval estimate of x
    StatesEroExmin(i,TIME-mem1)=Ex-Ero;
    % Upper boundary of interval estimate of x
    StatesEroExmax(i,TIME-mem1)=Ex+Ero;
end
% Update of prior information
LsuMem{i}.lbox0=Ex-Ero;
LsuMem{i}.ubx0=Ex+Ero;
end

```

Appendix 4: Matlab Code for SIS Particle Filtering

The core of a code for the SIS has the following structure. It incorporates an inequality constraint and solves a degeneracy problem by finding an effective sample size and resampling.

```

x = zeros(n,m);
W_pred = 1/N*ones(1,N);
for t = 1:m
    disp(['Running for time: ', num2str(t)]);
    x_pred = zeros(n,N);
    e = zeros(length(hdata(:,1)),N);
    for i = 1:N % particles

```

```

% Step 1. Drawing samples from the importance function
x_pred(:,i) = T*x0(:,i) + Q*randn(n,1);
while x_pred(9,i) < 0 % constraint
    x_pred(:,i) = T*x0(:,i) + Q*randn(n,1);
end
% Step 2. Updating the weight
e(:,i) = hdata(:,t)-Z*x_pred(:,i);
% Find the likelihood value, normalize and update the weight
log_lik = exp(-0.5.*diag(e'*inv(R.^ 2)*e))';
W = log_lik./sum(log_lik);
W_pred = (W_pred.*W)./sum(W_pred.*W);
end
% Step 3. Deciding if resampling is necessary
N_eff = inv(sum(W_pred.^ 2)); % estimating effective sample size
% Resampling when necessary
if N_eff < N_threshold % resampling is running
    W_tmp = cumsum(W_pred);
    rndn_u = rand(N,1); % uniform random pick
    smcRSI = zeros(1,N); % index for a chosen sample
    for i = 1:N
        smcRSI(i) = min(find(rndn_u(i)-W_tmp<0));
    end
    x0 = x_pred(:,smcRSI);
    W_pred = 1/N*ones(1,N);
else % resampling is not necessary
    x0 = x_pred;
end
% Calculate states
for i = 1:n
    x(i,t) = sum(W_pred.*x0(i,:));
end
end
end

```