Robust Kalman Filter Using Robust Cost Function

Pradeep Kumar Rajput

Roll no. 213EC6267

Department of Electronics and Communication Engineering

National Institute of Technology, Rourkela

Rourkela, Odisha, India

June, 2015

Robust Kalman Filter Using Robust Cost Function

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Technology

in

Signal and Image Processing

by

Pradeep Kumar Rajput

Roll no. 213EC6267

under the guidance of

Prof. Upendra Kumar Sahoo

Department of Electronics and Communication Engineering

National Institute of Technology, Rourkela

Rourkela, Odisha, India

June, 2015

dedicated to my parents...

National Institute of Technology
Rourkela

CERTIFICATE

This is to certify that the work in the thesis entitled **”Robust Kalman Filter Using Robust Cost Function”** submitted by ***Pradeep Kumar Rajput*** is a record of an original research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of Master of Technology in Electronics and Communication Engineering (Signal and Image Processing), National Institute of Technology, Rourkela. Neither this thesis nor any part of it, to the best of my knowledge, has been submitted for any degree or academic award elsewhere.

Prof. Upendra Kumar Sahoo
Assistant Professor
Department of ECE
National Institute of Technology
Rourkela

National Institute of Technology
Rourkela

DECLARATION

I certify that

1. The work contained in this thesis is originally done by Mital A. Gandhi and I have implemented and verify the result under the supervision of my supervisor.

2. I have followed the guidelines provided by the Institute in writing the thesis.

3. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

4. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Pradeep Kumar Rajput

Acknowledgment

This work is one of the most important achievements of my career. Completion of my project would not have been possible without the help of many people, who have constantly helped me with their full support for which I am highly thankful to them.

First of all, I would like to express my gratitude to my supervisor **Prof. Upendra Kumar Sahoo**, who has been the guiding force behind this work. I want to thank him for giving me the opportunity to work under him. He is not only a good Professor with deep vision but also a very kind person. I consider it my good fortune to have got an opportunity to work with such a wonderful person.

I am also very obliged to **Prof. K.K. Mahapatra**, HOD, Department of Electronics and Communication Engineering for creating an environment of study and research. I am also thankful to Prof. A.K Sahoo, Prof. L.P. Roy and Prof. S. Ari for helping me how to learn. They have been great sources of inspiration.

I would like to thank all faculty members and staff of the ECE Department for their sympathetic cooperation. I would also like to make a special mention of the selfless support and guidance I received from PhD Scholar Mr. Sanand Kumar and Mr. Nihar Ranjan Panda during my project work.

When I look back at my accomplishments in life, I can see a clear trace of my family's concerns and devotion everywhere. My dearest mother, whom I owe everything I have achieved and whatever I have become; my beloved father, who always believed in me and inspired me to dream big even at the toughest moments of my life; and my brother & sister; who were always my silent support during all the hardships of this endeavor and beyond.

Pradeep Kumar Rajput

Abstract

Kalman filter is one of the best filter used in the state estimation based on optimality criteria using system model and observation model. A common assumption used in estimation by Kalman filter is that noise is Gaussian but in practically we get thick-tailed non-gaussian noise distribution.This type of noise known as Outlier in the data and cause the significant degradation of performance of Kalman Filter.

There are many Nonlinear methods exist which can give desired estimation in the presence of Non-Gaussian Noise. We also want the filter which is Robust in the presence of outlier and statistically efficient.But classical Kalman Filter is not suitable in the presence of non-gaussian noise. To get the high statistical efficiency in the presence of outliers, A new robust Kalman Filter is used which can suppress observation, innovation and structural outlier by applying a new type of estimator , known as generalized maximum likelihood estimator.

This disquisition also contains the solution of different type of Nonlinear model that direct more than one equilibrium points.it is highly desirable to track the transition of state from one equilibrium point to another equilibrium point.This tracking method is computationally simple and accurate and also follow rapid transition effectively.

By simulations, the performance of GM-kF to different outliers and state estimation for the given applications: tracking of the vehicle and tracking climate transitions.

Keywords: Outliers, NonGaussian, Maximum likelihood, Robust,Thick-tailed.

Contents

List of Figures

List of Acronyms

Acronym	Description
AO	Additive Outlier
IO	innovation outlier
KF	Kalman Filter
MV	Minimum variance
LIT	Linear Invertible Transformation
MAP	Maximum a Posteriori
IRLS	Iterative recursive least square
EKF	Extended kalman filter
MLE	Maximum likelihood estimator
MAD	Median absolute deviation
GPS	Global positioning system
PS	Projection statistics
BP	Breakdown point
SD	standard deviation

Chapter 1

Introduction

Type of Outlier

Literature review

Reasearch objective

Result and Application

Chapter 1
Introduction

1.1 Introduction

The primary objective of Adaptive filter is used in signal processing is to extract noise & outliers from the signal of interest. The actual system can be represented via static or dynamic equation by discrete or continuous model. In general we consider system contaminated by Gaussian noise but practically this noise may be thick-tailed or non-Gaussian which introduce outliers in the system that is known as innovation and observation outliers[14, 13].

The classical Kaman filter requires the exact knowledge of noise distribution of the system and is not able to remove or overcome the outlier from the given system. When noise is non-Gaussian then it is very hard to suppress from system so we require a robust estimation that should be highly efficient in terms of statical efficiency and it should contain multiple outliers which occurring simultaneously in system.

1.2 Type of Outliers

To define the types of outlier discrete dynamic system is considered contaminated with Gaussian noise having the noise component with unknown distribution or thick-tailed.This type of contamination in the signal can cause the biased estimation or breakdown[19, 4]. The impulsive noise is present in many areas such as in the radio signal processing and radar the electromagnetic and acoustic interference from natural and man-made source. In indoor wireless

Types of noises	**Names of outliers in this work**	**Name of Outlier in the Literature**	**Affected model components**
Observation noise,e_k	Observation outlier	Isolated,Type I, Additive	z_k
system process noise ,w_k and control vector,u_k	Innovation outlier	Patchy, Type II, Innovation	$\hat{x}_{k\|k-1}$
Structural errors in H_k and F_k	structural outlier	-	z_k, $\hat{x}_{k\|k-1}$, $P_{k\|k-1}$, $P_{k\|k}$

communication the noise produced by the microwave ovens or electromechanical switches present In the printer, elevator and copying machine. The biomedical sensor uses to study and monitor the brain activity such as MRI also have a non-Gaussian distribution of noise because of interference with the complex tissue present in the brain.In Gps navigation non-line of sight (NLOS) signal propagation, due to the obstacle such as trees or building, cause outlier in the measurement. On the computer different component such as peripheral component interconnect(PCI) bus, liquid crystal display (LCD) produce impulsive interference that degrade the performance of the system.

Generally, this type of outliers can be seen in the time series analysis, linear regression model and survey data with identically independent data[18].We consider first two in this work. A mathematically unique and definite definition is not given in the literature. But Barnett and Lewis [10] defined them as a patch of observations which appears to be inconsistent with the remainder of that set of data. So by above definition we can say that the outlier is a set of data which does not follow the distribution followed by the majority of data[1]. Because this is the data which is generated by the other mechanism not by the system that produce the rest of data.

Martin and Yohai[15] shows two type of outlier namely isolated and patchy outlier. The outlier effects directly through the observation noise to the observation vector known as type I outlier .

$$z_t = H_t x_t + e_t \tag{1.1}$$

Another type of outlier i.e. patchy type affects the system state x_t of system dynamic model by w_t it is known as type II outlier.

$$x_t = F_t x_{t-1} + w_t + u_t \tag{1.2}$$

In field of engineering these two type of outlier also called as additive outlier and innovation outlier. The former one is also called as observation outlier. As a result, unique observations will be erroneously very large, but note that only one observation is affected, as the error does not enter the state of the model[19]. An example for this can be seen in satellite navigation. The state of the system is the dimensions of the satellite in space, whereas ground control receives a possibly noisy signal about this state. If there is short defect in the measurement device then it could cause AO. So the task of the estimator is to down-weight the influence of large observations for the state estimation. for this purpose robust Kalman filter are designed .IO outliers are carried to next step through out the observation.The task of estimator is to detect the outlier & adapt new condition as soon as possible. There is another type of outlier present in the system model known as structural outlier, this outlier affects the z_t and x_t of the observation and system model through wrong data in the matrices F_t and H_t .

In linear regression there is two type of outlier is defined vertical outlier and bad leverage point. These outlier may cause biased result or measurement fault in the system. Vertical outlier is data whose projection is not outlying in the model where majority of data falls and the bad leverage point is specifically a data which is far from the projection. We can show that later one can cause the severe effect in the maximum likelihood type estimator[9]. So we can treat innovation and observation outlier as vertical outlier and structural outlier as bad leverage point.

1.3 Literature Review

We now discuss classical and modern filtering technique used in continuous and discrete model of the system to find the best estimation of the system at each time step. Two linear estimator namely kalman filter and Luenberger estimator. The later one is use in the system model having deterministic noise[2]. In this method system stability maintain by correcting the system state by amount proportional to prediction error . some times system model is unknown in such cases we need to identify the system model and the prediction error may be due to system modification or state estimation. If system model parameter is known with additive system process and observation noise the most popular technique attributed to R.E. kalman from early 1960s [14, 13].At each time step t, the state vector $x_t \in \Re^{n\times 1}$ represent the system dynamic via(1.1)and observation vector $z_t \in \Re^{m\times 1}$ observation model via (1.2). In this equations $e_t \in \Re^{m\times 1}$ represents the observation noise, $w_t \in \Re^{n\times 1}$ represents system process noise at time t , $u_t \in \Re^{n\times 1}$ represents input control matrix $F_t \in \Re^{m\times m}$ represents the state transition matrix, and $H_t \in \Re^{m\times n}$ represents the Observation matrix. Kalman filter follows two assumption (I) a system follows the Markov process i.e. process whose future behavior can be predicted accurately from current and future behavior and does not depend completely on its past behavior its past behavior and also involves random chance or probability. (II) the noise present in system and observation are consider as zero mean white Gaussian noise.i.e.

$$w_t \sim N[0, W_t] \tag{1.3}$$

$$e_t \sim N[0, R_t] \tag{1.4}$$

In the literature, many methods are given for a non-linear model with different kind of noise having Non-Gaussian Noise distribution that affect the system and observation process [18, 19].To handle non-Gaussian noise Bucy proposed one nonlinear filter [2], If order of state variable will increase then this method is computationally intensive and assume noise distribution is apriori.

To handle the different kind of outliers many methods are proposed namely Doblinger's adaptive Kalman method [13]. Durovic, Durgaprasad and Kovacevic [5] used M-estimator to remove outlier, but main problem in this approach was when solving the nonlinear estimator it does not recapitulate at each time step, assume that the all the prediction are accurate and remove the observation outlier that deviate from mean as result of that when innovation and observation outlier concurrently occurs it gives unreliable result.Hence, a filter is needed that suppress innovation and observation outlier simultaneously and does not depend wholly on the observation or prediction.The covariance matrix obtained by classical Kalman Filter is inaccurate in this method. The only oddity in the method proposed by Durovic and Kovacevic [5] and huber[10] is the covariance matrix of M-estimator.

The other method that used mostly for speech signal enhancement is moving median filter . But for this method to work the filter window should be twice as long as falsify sequence.So filter submits deteriorated estimate when outlier occurs simultaneously in several samples.Another method proposed to detect outlier and suppress them by Mital A. Gandhi [1] can contain simultaneously occur outlier using generalized maximum likelihood estimator. If noise is unknown but bounded then, H_∞ filter is used that can robustify the modelling error.

1.4 Research Objective

In literature, not much method is available to suppress and detect the innovation, observation and structural outlier simultaneously. Effect of observation outlier is, it affects the single observation and return to the normal path, innovation outlier affect the set of data and carried out to the next stage so the estimation of state may be completely biased, So we must remove these outliers. The method described in section 2.3 is not able to handle all outlier simultaneously and yield an inaccurate result.Outlier may defined as the data which is far from data cloud. We don't know the source that generate the outlier in the system hence using maximum likelihood estimator optimal estimator can not designed and can not be used because it is not robust to the structural outlier.

The objective of the research is to develop an estimator that should be robust and can handle all the three types of outliers with positive breakdown point. Breakdown point shows the most extreme part of exception (highly deviating samples)that an estimator can deal without breaking down. The filter not only should be robust it also should be a good estimator in classical statistical terms that identified by properties of consistency, unbiasedness, rate of convergence, and efficiency. First, for a good estimator rate of convergence towards the actual value should be fast.Second, it should follow the Fisher consistency i.e. when number of measurement increases the estimator should converge towards the correct value of the parameter to be estimated.Third, the estimator should be unbiased, i.e. the actual value for any sample size its mean value should be equal.Fourth, the variance of the estimates follow the Cramer-Rao bound. In summary, we are interested in robust and highly efficient filter means it should have positive breakdown point and continues to keep good performance for Gaussian noise observation.

1.5 Application and Results

The Kalman Filter has a large number of application in science and technology. one common application in spacecraft and aircraft for guidance, control and navigation of vehicle. Also used in statistical signal processing for time series analysis. In field of robotic motion & control kalman filter is main topic ,sometimes they also include trajectory optimization.

Chapter 2

Introduction

Chapter 2

Analysis of Classical Kalman Filter & other Filtering Technique

The KF suppress noise by considering a predefined model of system.Therefore modelling of KF should be correct &meaningful. It should be define as follow:

1. **Understand the situation** : Take a look at the issue break it down to the scientific fundamental.On the off chance if we will not do this it may lead to unnecessary work[3].

2. **Model the state process & measurement process** : Start with the basic modelling of state this model may not work correctly & effectively at first but this can be refined later.Also analyze how we can measure the process.The measurement space may not be same as state space.

3. **Noise modelling** : the assumption made at the time of development of kalman filter is noise should Gaussian . so this should be done for both the state & measurement process[14].

4. **Understand the situation** : Now test that filter is behaving correctly or not ,if not then use synthetic data & also try to change noise parameter.

Kalman Filter can be visualize an estimator that takes noisy measurement sequence & produces three types of output based on the associated model[3].

- **State estimator or Reconstructor** : It process noisy measurement $y(t)$ & build state estimate $x(t)$.

- **Measurement filter** : In this step it produces a filtered output measurement sequence $\{\hat{y}(t|t)\}$by accepting noisy sequence in $\{y(t)\}$ in input.
- **Whitening filter** : Produces white measurement $\{e(t)\}$ or uncorrelated sequence by processing noisy input sequence $\{y(t)\}$ which is correlated . This uncorrelated sequence is also known as innovation sequence.

2.1 Kalman Filter Algorithm

Kalman filter algorithm can be derived from the innovation point of view following the approach by Kailath[12, 3].State space model of a stochastic process can be represent by equation:

$$x_t = F_t x_{t-1} + w_t + u_t \tag{2.1}$$

where w_t is assumed zero mean white Gaussian noise with covariance R_{ww} and x_t & w_t are uncorrelated. The measurement model is represented by equation :

$$z_t = H_t x_t + v_t \tag{2.2}$$

where v_t is a zero mean white Gaussian noise with covariance R_{ww} and v_t is uncorrelated with x_t and w_t. Limiting the estimator to be linear[12],the MV estimator for a batch of N data is represented by:

$$\hat{X}_{MV} = K_{MV} Z = R_{xz} R_{zz}^{-1} Z \tag{2.3}$$

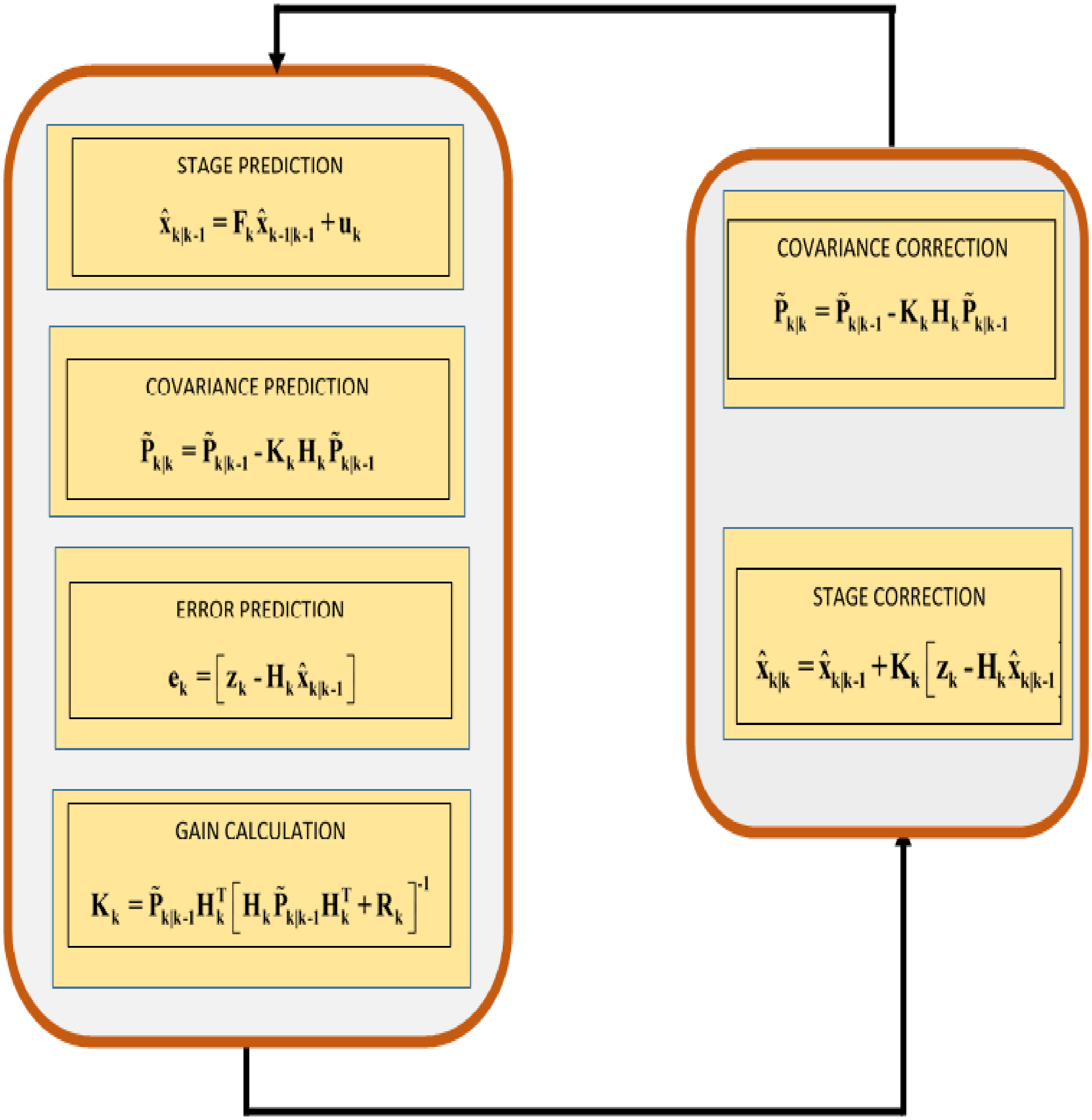
where $\hat{X}_{MV} \in R^{N_x N \times 1}$,R_{xz} and $\hat{X}_{MV} \in R^{N_x N \times N_z N}$,$R_{zz} \in R^{N_z N \times N_x N}$,and $Z \in R^{N_z N \times 1}$. Representing this by set of N data sample:

$$\hat{X}_{MV}(N) = K_{MV}(N) Z(N) = R_{xz}(N) R_{zz}^{-1}(N) Z(N) \tag{2.4}$$

where $\hat{X}'_{MV}(N) = [\hat{x}'(1).........\hat{x}'(N)]'$, $Z(N) = [z'(1)...........z'(N)]'$, $\hat{x} \in R^{N_x \times 1}$, and $z \in R^{N_z \times 1}$. For a "batch"solution of all state estimation problem we process all N_z-vector data $\{z(1)........z(N)\}$ in one batch. A recursive solution to this problem in form of

$$\hat{X}_{new} = \hat{X}_{MV} + K \ E_{new} \tag{2.5}$$

Transform the covariance matrix R_{zz} in block diagonal to achieve recursive


STAGE PREDICTION
$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + u_k$
COVARIANCE PREDICTION
$\tilde{P}_{k|k} = \tilde{P}_{k|k-1} - K_k H_k \tilde{P}_{k|k-1}$
ERROR PREDICTION
$e_k = \left[z_k - H_k \hat{x}_{k|k-1} \right]$
GAIN CALCULATION
$K_k = \tilde{P}_{k|k-1} H_k^T \left[H_k \tilde{P}_{k|k-1} H_k^T + R_k \right]^{-1}$
COVARIANCE CORRECTION
$\tilde{P}_{k|k} = \tilde{P}_{k|k-1} - K_k H_k \tilde{P}_{k|k-1}$
STAGE CORRECTION
$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left[z_k - H_k \hat{x}_{k|k-1} \right.$

solution

$$R_{zz}(N) = \begin{bmatrix} E\{z(1)z'(1)\} & . & . & . & E\{z(1)z'(1)\} \\ . & & . & & . \\ . & & . & & . \\ . & & . & & . \\ E\{z(N)z'(1)\} & . & . & . & E\{z(N)z'(N)\} \end{bmatrix}$$

$$R_{zz}(N) = \begin{bmatrix} R_{zz}(1,1) & . & . & R_{zz}(1,N) \\ . & & . & . \\ . & & . & . \\ R_{zz}(N,1) & . & . & R_{zz}(N,N) \end{bmatrix}$$

Block diagonal of $R_{zz}(N)$ implies that off diagonal elements of matrices $R_{zz}(i,j) = 0$ for $i \neq j$,which implies that $\{z_t\}$ must be orthogonal (uncorrelated). so new independent sequence of N_z vectors, say $\{e_t\}$, such that

$$E\left\{e_t e_k'\right\} = 0 \qquad for \qquad t \neq k \tag{2.6}$$

Now innovation can be defined as

$$e_t := z_t - \hat{z}_{t|t-1} \tag{2.7}$$

innovation sequence follows the orthogonality property that

$$cov[z_T, e_t] = 0 \qquad for \qquad T \leq t-1 \tag{2.8}$$

$\{e_t\}$ is a time uncorrelated measurement vector so we have

$$R_{ee}(N) = \begin{bmatrix} R_{ee}(1) & & 0 \\ & . & \\ & & . \\ 0 & & R_{ee}(N) \end{bmatrix} \qquad for \;\; each \quad R_{ee}(i) \in R^{N_z \times N_z}$$

If the measurement vector is correlated then through linear transformation it can transformed into uncorrelated innovation vector[18],say $\mathbf{S}$ given by

$$Z(N) = \mathbf{S}e(N) \tag{2.9}$$

where $\mathbf{S} \in R^{N_zN \times N_zN}$ is nonsingular matrix and $e := [e'(1)\,............e'(N)]'$, First multiply Eq.(3.9) by its transpose and take expected value of solution obtained, we get

$$R_{zz}(N) = \mathbf{S}R_{ee}(N)\mathbf{S}'$$

Now taking inverse of above equation, we get

$$R_{zz}^{-1}(N) = (\mathbf{S}')^{-1}R_{ee}^{-1}(N)\,\mathbf{S}^{-1}$$

similarly we obtain

$$R_{xz}(N) = R_{xe}(N)\mathbf{S}'$$

$$\hat{X}_{MV}(N) = K_{MV}(N)\,Z(N) = R_{xz}(N)\,R_{zz}^{-1}(N)\,Z(N) = [R_{xe}(N)\mathbf{S}'][(\mathbf{S}')^{-1}R_{ee}^{-1}(N)\,\mathbf{S}^{-1}]Se(N)$$

or

$$\hat{X}_{MV}(N) = R_{xe}(N)R_{ee}^{-1}(N)\,e(N) \tag{2.10}$$

we already know that e_t is orthogonal so it can be shown that $R_{ee}(N)$ is lower block triangular

$$R_{xe}(N) = \begin{cases} R_{xe}(t,i), & t > i \\ R_{xe}(t,i), & t = i \\ 0, & t < i \end{cases}$$

substituting into Eq.(5.10) we get

$$\hat{X}_{MV}(N) = \begin{bmatrix} R_{xe}(1,1) & . & . & R_{xe}(1,N) \\ . & . & & . \\ . & & . & . \\ R_{xe}(N,1) & . & . & R_{xe}(N,N) \end{bmatrix} \begin{bmatrix} R_{ee}^{-1}(1) & . & . & 0 \\ . & . & & . \\ . & & . & . \\ 0 & . & . & R_{ee}^{-1}(N) \end{bmatrix} \begin{bmatrix} e(1) \\ . \\ . \\ e(N) \end{bmatrix} \tag{2.11}$$

For best estimate of x_t given Z_t Eq.(5.11) can be written as(where N=t)

$$\hat{x}_{t|t} = \sum_{i=1}^{t} R_{xe}(t,i)R_{ee}^{-1}(i)\,e(i)$$

Extracting last (t^{th}) term from above sum we get

$$\hat{x}_{new} = \hat{x}_{t|t} = \underbrace{\sum_{i=1}^{t-1} R_{xe}(t,i)R_{ee}^{-1}(i)\,e(i)}_{\hat{x}_{old}} + \underbrace{R_{xe}(t,t)R_{ee}^{-1}(t)\,e(t)}_{K} \tag{2.12}$$

or

$$\hat{x}_{new} = \hat{x}_{t|t} = \hat{x}_{t|t-1} + Ke_t \tag{2.13}$$

minimum variance estimate of z_t is same as minimum variance estimate of x_t;

given by

$$\hat{z}_{t|t-1} = H_t \hat{x}_{t|t-1} \tag{2.14}$$

so innovation can be solved using equations (5.2) and (5.14) as

$$e_t = z_t - H_t \hat{x}_{t|t-1} = H_t \left[x_t - \hat{x}_{t|t-1} \right] + v_t$$

properties of innovation sequence summarize as follows:

1.innovation sequencee_t is white & Gaussian under the Gauss-Markov assumption with distribution $\mathrm{N}\left(0, R_{ee}(t)\right)$.

2.innovation sequence e_t is zero mean.

3.innovation sequence e_t is uncorrelated in time and with input u_{t-1}.

4.Innovation sequence e_t and measurement z_t are equivalent under LIT.

$$e_t = H_t \tilde{x}_{t|t-1} + v_t \tag{2.15}$$

$\tilde{x}_{t|t-1} = x_t - \hat{x}_{t|t-1}$ is prediction stae estimation error . using above equation innovation covariance is given as

$$R_{ee}(t) = H_t \tilde{P}_{t|t-1} H_t' + R_{vv}(t) \tag{2.16}$$

here v and $\tilde{x}$ is uncorrelated. The gain matrix is given as

$$K = R_{xe}(t) R_{ee}^{-1}(t) = \tilde{P}_{t|t-1} H'_t R_{ee}^{-1}(t) \tag{2.17}$$

The predicted error covariance $\tilde{P}_{t|t-1} = \mathrm{cov}(\tilde{x}_{t|t-1})$ is

$$\tilde{P}_{t|t-1} = F_{t-1} E\{\tilde{x}_{t-1|t-1} \tilde{x}'_{t-1|t-1}\} F'_{t-1} + E\{w_{t-1} \tilde{x}'_{t-1|t-1}\} F'_{t-1} + F_{t-1} E\{\tilde{x}_{t-1|t-1} w'_{t-1}\} + E\{w_{t-1} v \tag{2.18}$$

w and $\tilde{x}$ are uncorrelated so we get

$$\tilde{P}_{t|t-1} = F_{t-1} \tilde{P}_{t-1|t-1} F'_{t-1} + R_{ww}$$

The corrected error covariance $\tilde{P}_{t|t}$ calculated using corrected state estimation error and corresponding state estimate of Eq.(3.13)

$$\tilde{x}_{t|t} := x_t - \hat{x}_{t|t} = x_t - \hat{x}_{t|t-1} - K e_t$$

or

$$\tilde{x}_{t|t} = \tilde{x}_{t|t-1} - K e_t \tag{2.19}$$

From above we calculate the error covariance as

$$\tilde{P}_{t|t} = [I - KH_t]\tilde{P}_{t|t-1} - \tilde{P}_{t|t-1}H'_tK' + \tilde{P}_{t|t-1}H'_tR_{ee}^{-1}(t)R_{ee}(t)K' \tag{2.20}$$

solving the Eq.(3.17) and (3.20) we get corrected error covariance as

$$\tilde{P}_{t|t} = [I - KH_t]\tilde{P}_{t|t-1} \tag{2.21}$$

Some of the important characteristics of Kalman filter is discussed here[3].It is MAP estimator,an unbiased and recursive Bayesian.The filter estimate is an ML-estimate When noise is Gaussian, and the MSE is the minimum criterion. Even if an assumption of Gaussian noise does not satisfythe condition filter is best MV estimation filter,i.e. for all class of linear filters it minimizes the variance of estimation error. Due to its recursive and linear behaviour KF is computationally fast. The a priori error covariance can be calculated offline that does not depend on the actual error. But Kalman filter has some of the Non negligible drawbacks[9].First, if covariance matrix's estimation is not accurate, the optimal performance can not be achieved. Also, if noise process follows Non-Gaussian distribution or system model parameter have structural outlier then also optimal performance can not be achieved. second, the filter completely trust on the observations and predictions if the solution is not iterating, Due to this the filter is biased even if the single outlier is present because of this outlier any deviations or errors from the assumption can not be seized.If error is unmodeled then filter can be make robust by increasing noise covariance matrix w_t which increase the Kalman gain matrix **K** through the $\tilde{P}_{t|t-1}$.

2.2 Application of Classical Kalman Filter in RC tuning circuit

In this example we measure the voltage of an RC circuit using a voltmeter,having high impedance[3]. This measurement is contaminated with random noise that can be modeled by

where

V_{out} = calculated voltage

K = instrument amplification factor

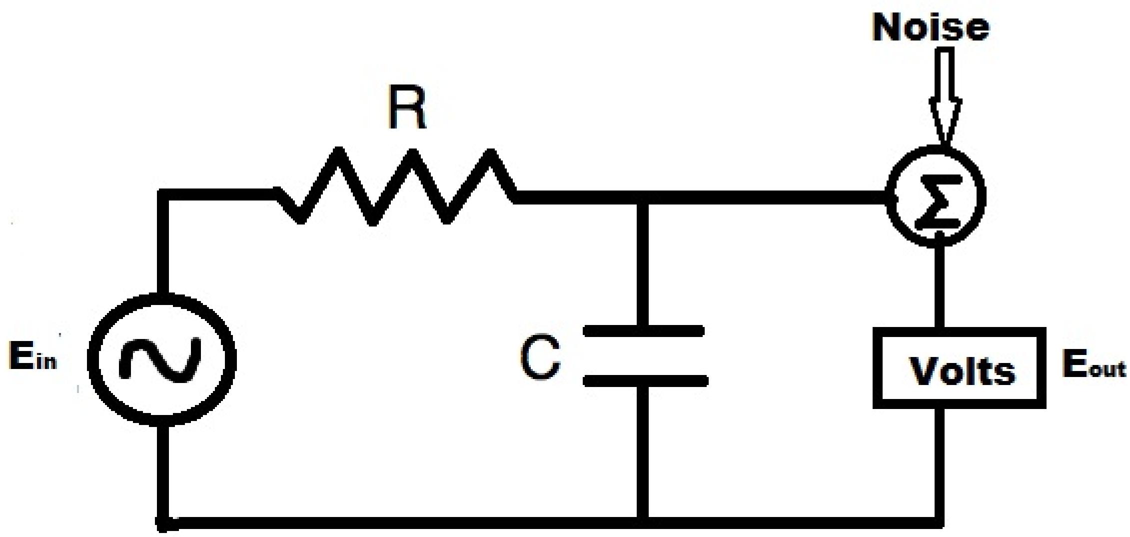


Figure 2.2: RC circuit model

V = true voltage

η = random noise with zero mean with noise variance of $R_{\eta\eta}$

Solving the above circuit by applying Kirchhoff's current law at the node we get,

$$I - \frac{Vin}{R} - C\frac{dV_{in}}{dt} = 0$$

where initial voltage is given by V_0 and R is resistance, C is capacitance. For a voltmeter gain K measurement equation given by:

$$V_{out} = KV_{in}$$

taking first difference of above equation :

$$C\frac{Vin(t) - Vin(t-1)}{\Delta T} = -\frac{Vin(t-1)}{R} + I(t-1)$$

or

$$Vin(t) = \left(1 - \frac{\Delta T}{RC}\right) Vin(t-1) + \frac{\Delta T}{C} I(t-1)$$

for above circuit consider $R = 3K\Omega$ and $C = 1000\mu F$, $V_0 = 2.5V$, $\Delta T = 100ms$, $K = 2$ .Now transform physical model into state space model by considering

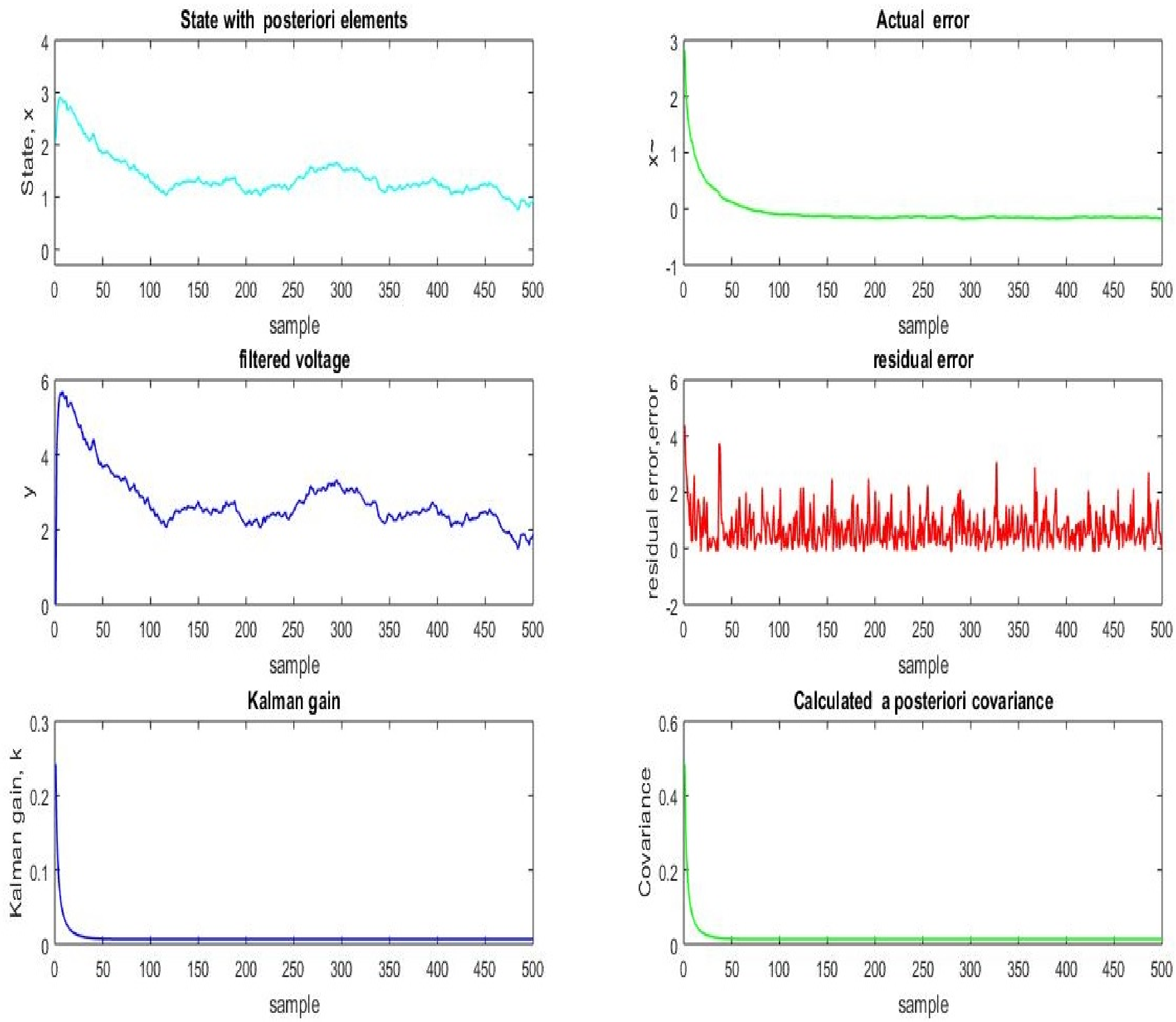


Figure 2.3: Result Tuning of RC circuit

$x = V_{out}$ and $u = I$ we get the equation:

$$x(t) = 0.97x(t-1) + \eta(t-1) + 100u(t-1)$$
$$y(t) = 2x(t) + v(t)$$

Chapter 3

Robust estimation

Properties of estimators

Outliers,Leverage point & influential point in regression

Properties of Robust estimator

Maximum Likelihood estimation

Types of Robust estimator

Chapter 3
Robust Estimation

3.1 Properties of estimators

The properties of estimator is discussed from classical and robust point of view for estimation of scale, location and scatter[19]. Let $X = \{x_1, x_2, ...x_n\}$ is a set of n i.i.d random data , which satisfy the univariate model $x = z + e$.

3.1.1 scale estimators

scale estimators provide estimate of span around the location of sample. Some location free estimators are also available. Standard deviation is measure of scale , **SD** is square root of of the variance σ^2 & denoted by σ .

$$\sigma^2 = E\left[(x-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx \tag{3.1}$$

the variance in discrete case is given by

$$\sigma^2 = \sum_{i=1}^{m} (x_i - \mu)^2 p(x_i) \tag{3.2}$$

then MLE of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{3.3}$$

But estimator is biased for n i.i.d samples ,to make it unbiased replace n in equ. (5.3) by 1/(n-1). For some distribution e.g. Cauchy distribution S.D. does not exist.For univariate sample of quantitative data median absolute deviation (MAD) is a robust measure of the variability and defined as

$$\mathbf{MAD} = 1.4826k \, \underset{i}{med} \left| x_i - \underset{j}{med}(x_j) \right| \tag{3.4}$$

Where k is correction factor and given by

$$k = \frac{m}{m-8} \tag{3.5}$$

Because of this estimation is unbiased and satisfy the criteria of Fisher consistency[7].

3.1.2 location estimators

Estimation of location can be done by the sample mean parameter. A LSE which is maximum likelihood under Gaussian distribution and minimizes the loss function

$$\xi(x) = \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} (x_i - z)^2 \tag{3.6}$$

It estimate the expected value of random process, also known as mean , defined as measure of the central tendency either of a probability distribution.

$$\mu = E[x] = \int_{-\infty}^{+\infty} x f(x)\, dx \tag{3.7}$$

Expected value of mean for discrete case is given as

$$\mu = E[x] = \sum_{i=1}^{n} x_i p(x_i) \tag{3.8}$$

and sample mean i.e. estimaed value of mean is given by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3.9}$$

If single outlier occur in the sample data then sample mean estimator lead to biased estimate which cause the breakdown point zero[19]. discussed in previous chapter breakdown point is measure of maximum number of outlier that an estimator can handle effectively.Median is another estimator used to measure location and its center of probability is defined as

$$\int_{-\infty}^{med} f(x)\, dx = \int_{med}^{+\infty} f(x)\, dx = \frac{1}{2} \tag{3.10}$$

For an ordered sequence of data sample median is defined as middle value. For this sort data in increasing or decreasing order then find the value that is halfway through this ordered sequence. If number of data is odd then median will be middle value.Let $\upsilon = [m/2] + 1$ where [.] represents integer part.Further sample median given by

$$\hat{x}_{med} = \begin{cases} x_{\upsilon}, & for\, n\, odd \\ (x_{\upsilon-1} + x_{\upsilon})/2, & for\, n\, even \end{cases} \tag{3.11}$$

3.1.3 scatter estimator

For multi-dimension i.e. multivariate data , the scatter of sample is measured on the basis of correlation information obtained from covariance matrix [5]. As example , an $n \times n$ covariance matrix of a vector x with zero mean for n observation of length $n \times 1$ is given by

$$R = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \tag{3.12}$$

For one dimensional case sample mean and sample variance is used, in the presence of outlier it is liable to breakdown point.

3.2 Outliers, Leverage & Influential points in regression

- Outlier: An observation point that does not follow the distribution & distant from other observations. It may indicate experimental error or may be due to variability in the measurement; the former are sometimes excluded from the data set[15].

- Influential point: They are outliers, i.e. graphically they are far from the pattern described by the other points, that means that the relationship between x and y is different for that point than for the other points. Observations with very low or very high value of x are in positions of high leverage .

- Leverage points: are those observations, if any, having lack of neighboring observations that the fitted regression model will pass close to that particular observation. Outliers that are not in a high leverage points or high leverage position that are not outliers and do not tend to be influential called leverage point.
  par Effect of this outlier shown in least square estimator.

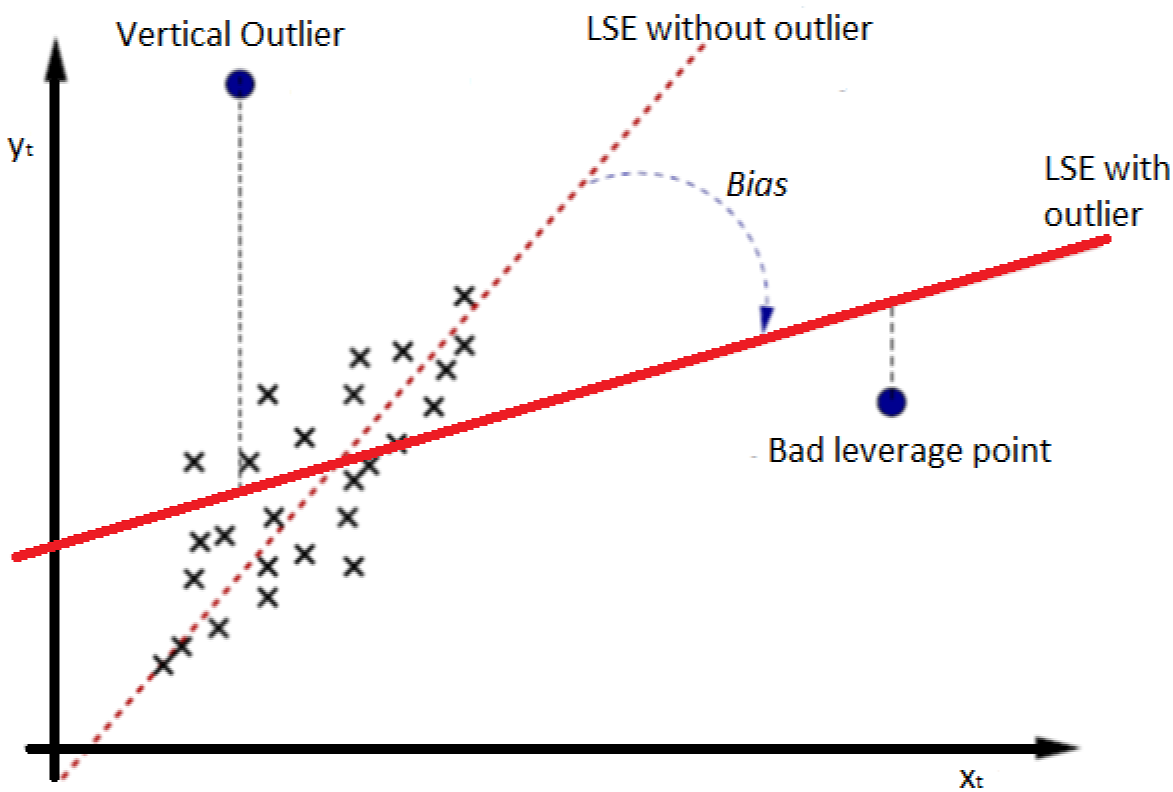


Figure 3.1: Effect of outlier & bad leverage point in linear regression

3.3 Properties of Robust estimator

To work in signal processing need to understand some concept of robust statistics, which we return to in this segment'

3.3.1 Quantitative robustness

In an estimator, quantitative robustness is characterized by BP. The breakdown point defined as maximum number of highly deviating observation that can be handled by an estimator. It can take value in the range 0-50% . Larger quantitative robustness achieved when breakdown point is high. For a sample mean breakdown point is zero that means a single outlier may degrade estimator performance completely[19]. But in case of sample median BP is 50% . after 50% it is hard to discriminate between the contamination and normal distribution.

3.3.2 Qualitative robustness

: It is characterized by influence function. In case of extremely small contamination bias effect at an arbitrary point is described by IF. The asymptotic IF is defined ,when limit exist, for an estimator $\hat{\theta}$and nominal distribution H_θ by first derivative of its functional version as

$$IF\left(x;\hat{\theta},H_\theta\right) = \lim_{\varepsilon\to 0} \frac{\hat{\theta}_\infty(H) - \hat{\theta}_\infty(F_\theta)}{\in} = \left[\frac{\partial \hat{\theta}_\infty(H)}{\partial \in}\right]_{\varepsilon\to 0} \tag{3.13}$$

when data is distributed then $\hat{\theta}_\infty(H)$ and $\hat{\theta}_\infty(F_\theta)$ are asymptotic value of estimator. Influence function satisfies the properties such as continuity and boundness. Former one ensures that the small change in data causes the little change in estimate.Whereas boundness make certain that the small fraction of outlier or contamination has slight effect on the guess. A filter is said qualitatively robust only when both properties satisfied.

3.4 Maximum Likelihood estimation

maximum likelihood estimator is often abbreviate by MLE. MLE is very important technique for estimating a parameter of distribution & sometimes may be it is simplest technique and often it gives most natural estimate.

setup : given data $D = \{x_1,x_2,........x_n\}$ $x :\leftarrow \Re^d$ assume a set of distribution $\{P_\theta,\theta \leftarrow \Phi\}$ $on\ \Re^d$ assume D is a sample from $\{x_1,x_2,........x_n\}$ is distributed according to one of P_θ is i.i.d for some $(\theta \in \Phi)$.

Goal: the goal of the estimator is to choose or estimate the true value of θ.

Definition:θ_{MLE} is a maxium likelihood estimate for θ is

$$\theta_{MLE} = \underset{\theta\in\Phi}{\arg\max}\, P(D|\theta)$$

Presicely it can be written as

$$P(D|\theta_{MLE}) = \max_{\theta\in\Phi} P(D|\theta)$$

where we can write probability in another way as

$$P_\theta(x) = P(x|\theta)$$

here,

$$P(D|\theta) = P(x_1, x_2, ...x_n|\theta)$$
$$\prod_{i=1}^{n} P(x_i|\theta) = \prod_{i=1}^{n} P(X_i = x_i|\theta)$$

Remark:

1. An MLE might not be unique.
2. MLE may fail to exist , there might not be a θ that achieves

$$P(D|\theta_{MLE}) = \max_{\theta \in \Phi} P(D|\theta)$$

Pros:

1.Easy to compute.

2.Often interpretable.

3.Having nice asymptotic property:

a) Consistent – as number of sample goes to ∞ or number of point increases it converges to true value of θ with high probability.

b) Efficient – It is the best possible estimate of true data having lowest variance i.e. low error.

c) Invariant under reparameterization – it means that if for any function $g(\theta_{MLE})$ is a MLE for $g(\theta)$ then if we square θ and want MLE of square then only need to square MLE of θ_{MLE} .

d) Normal – If N (number of sample) is very large then distribution looks like normal.

cons:

a) It is point estimation so no representation of uncertainities . Ideally it is not representative for spikes or impulsive noise in the likelihood function. b)wrong objective – It might maximize wrong objective disregarding loss function. c)Existence & uniqueness is not gauranteed.

Properties of MLE:

Unbiasedness:It tells about expectation of an estimator is true value of sample. Consistency: If number of sample increases then the value which estimated should tend to true value.

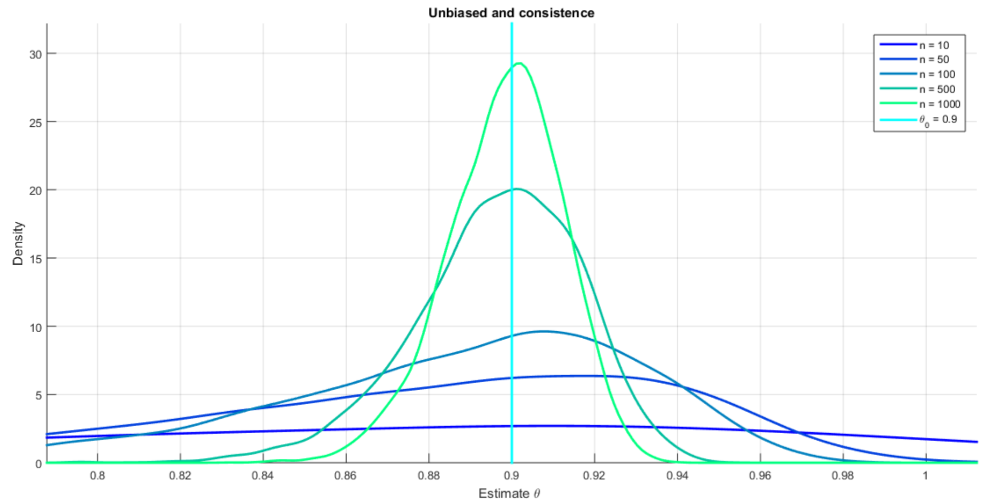


Figure 3.2: M-estimator property unbiased and consistence

3.5 Types of Robust estimator

if the data is deviated from the distributional assumption then we need robust statistic to suppress the outlier. In literature 3 type of Robust estimator is mentioned:

1.M-estimator

2.R-estimator

3.L-estimator In this work we mainly concentrate on M-estimator.It can resist outlier without preprocessing data.So this estimator is desribed in this work.

3.5.1 M estimator

In statistics, M-estimators are widely used estimator obtained as the minima of sums of functions of the data.One such example of m-estimator is Least-squares estimators.Based on the robust statistic different type of M-estimators are developed. The measurable technique of assessing an M-estimator on an information set is called M-estimation.In an M-estimator zero of an estimating function evaluated.Most of estimating function is obtained by derivatizing an additional statistical function: For example, an MLE is often achieved by derivatizing the likelihood function and find zero with respect to some parame-

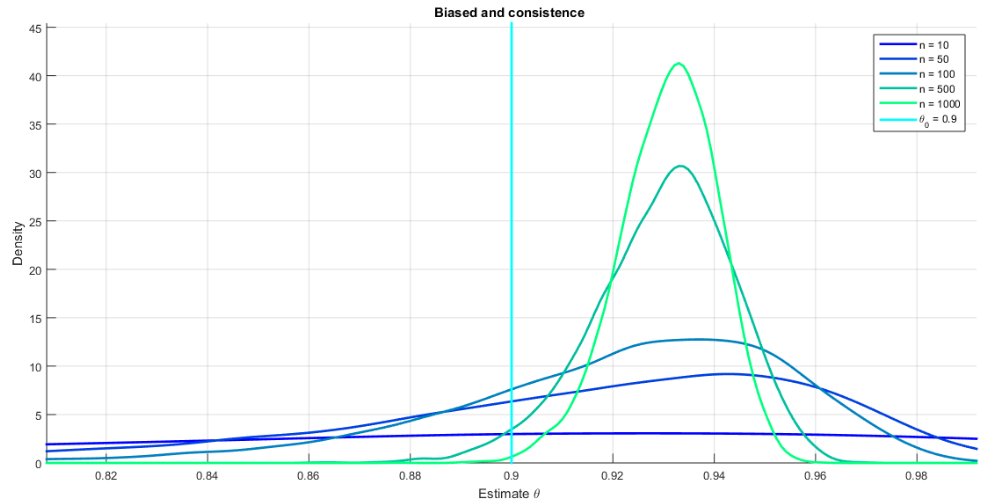


Figure 3.3: M-estimator property biased and consistence

ter.thus, an MLE is often a critical point of the score function.In single channel context this type of estimator is easily accessible & witout preprocessing data it can resist outliers.

M-estimator fits in withclass of maximum likelihood estimator. M-Estimatior minimizes thefollowing function:

$$\sum_{i=1}^{m} \xi\left(e_i\right) \tag{3.14}$$

where ξ is some function that satisfy following properties:

- $\xi\left(e\right) \geq 0$ for all e and has minimum at 0.
- $\xi\left(e\right) = \xi\left(-e\right)$ for all e.
- As e increases value of $\xi\left(e\right)$ also increases but its value does not get very large .

From the above properties some example of M-estimator is descibed here:

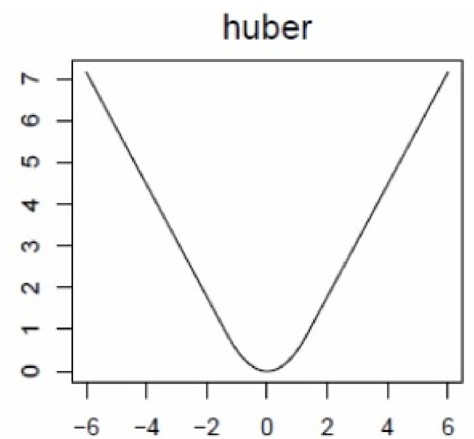
1. Huber's estimator
2. Bisquare weight estimator

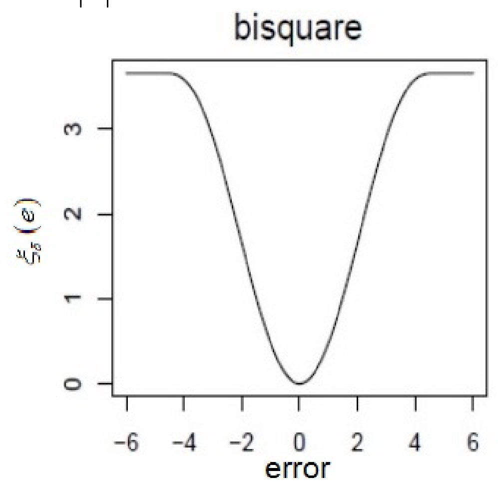
3.5.1.1 Huber Estimator

Huber lossis aloss function thatused in robust estimation. Squared error loss function is very sensitive to outlier, instead of this huber loss function is used that is less sensitive to outlier.The Huber loss function report the error sustained by anestimation approach. THe loss function defined piecewise as

$$\xi_\delta(e) = \begin{cases} \frac{1}{2}e^2 & for\ |e| \le \delta \\ \delta\left(|e| - \frac{1}{2}\delta\right), & otherwise \end{cases} \tag{3.15}$$

For small value of e this function is quadratic & for large values it is linear.The slope of two sections are equal at two points where $|e| = \delta$





3.5.1.2 Bi-square estimator

Bisquare estimator is alternative weighting scheme that weight the error (residual). From the unweighted fit we calculate residual & then apply weight function given below:

$$\xi_\delta(e) = \begin{cases} \frac{\delta^2}{6}\left\{1 - \left[1 - \left(\frac{e}{\delta}\right)^2\right]\right\} & for\ |e| \le \delta \\ \frac{\delta^2}{6} & for\ |e| > \delta \end{cases} \tag{3.16}$$

Chapter 4

Evolution of GM-Kalman Filter

Outlier, Breakdown point and statistical Efficiency

Necessity of Redundancy in GM-KF

GM-KF with redundancy & prewhitening

Projection statistic

Robust Filtering based on GM-estimation

Chapter 4
Evolution of GM-Kalman Filter

In this chapter, we discuss how to suppress outlier from the sample.Then we look for a compromise between the two statistical properties for robustness namely breakdown point and statistical efficiency. After that, we develop GM Kalman filter. We use batch mode regression in this filter to get desired redundancy.If an outlier is present in that sample, then a new pre-whitening method is used to decorrelate the data.In next step, we solve this by using GM-estimator for state estimation.

4.1 Outlier,Breakdown Point and Statistical Efficiency

There are some methods available in the literature that detect & discard the outlier. It is very easy & add no more complexity in estimation. But in some cases rejecting the erroneous data is not a suitable option it may lead to performance degradation of the estimator. So one alternative for such problem is down-weighting the outlier instead of deleting them.By the above method, statistical efficiency can maintained at some level especially for the outlier that is afar from the outlier detection threshold.

The breakdown point defined as maximum number of outlier that an estimator can handle. So if one can want to achieve high breakdown point for an estimator that they need to compromise with statistical efficiency, which is defined by Fisher. As already discussed that sample mean estimator has zero breakdown point , but if distribution is Gaussian then it is 100 % efficient, Later on sample median achieve highest breakdown point but its efficiency decreased, which is experimentally shown by Fisher efficiency it became 64 % for

Gaussian distribution. To properly address trade-off we consider huber function that has positive breakdown point of 35%.If we add some redundant data from the large data set, it will further increase the breakdown point.Projection statistic used to find the bad leverage point & down weight those points. From above discussion, it is clear that GM-KF will produce more accurate estimate in presence of outlier with good statistical efficiency.

4.2 Necessity of Redundancy in GM-KF

In GM-KF redundancy obtained by taking Batch-mode regression which gives positive breakdown point. An estimator's Breakdown point under common belief is represented by

$$\varepsilon^*_{\max} = [(m-n)/2]/m \tag{4.1}$$

where m is number of observation & n is number of state variables. In classical Kalman filter at each time step k , it collects n prediction i.e. one for each state variable and 1 observation so total m=n+1 observation available. so maximum Breakdown point is specified by

$$\begin{aligned} \varepsilon^*_{\max} &= [(n+1-n)/2]/m \\ &= [1/2]/m \\ &= 0/m = 0, \end{aligned} \tag{4.2}$$

i.e. zero Breakdown .To get positive breakdown point & make robust to outlier we consider two redundant measurement which gives total observation m=n+3.For this maximum Breakdown point denoted as

$$\begin{aligned} \varepsilon^*_{\max} &= [(n+3-n)/2]/m \\ &= [3/2]/m \\ &= 2/m, \end{aligned} \tag{4.3}$$

shows that estimator can now handle up to two outlier.so from above discussion it is clear that for every 2 observation estimator can handle one more outlier. Thus m/2 outlier can handled by adding $m_r = m-1$ redundant observation.

For example, to handle 4 outlier simultaneously i.e. $m/2 = 4$, total number of observation required by estimator is m=8 and additional measurement m_r

=7. Maximum breakdown point for this redundancy is

$$\begin{aligned}\varepsilon^{*}_{\max} &= [(n+8-n)/2]/m \\ &= [8/2]/m \\ &= 4/m,\end{aligned} \tag{4.4}$$

This type of observation redundancy can be achieved by batch-mode linear regression.

4.3 GM-KF with redundancy & prewhitening :

Discrete dynamic model of Observation & state space equation combine to achieve Batch mode regression. Observation equation is given by,

$$z_k = Hx_k + e_{k,} \tag{4.5}$$

prediction equation of state is represented as

$$\hat{x}_{k|k-1} = x_k + \delta_{k-1}, \tag{4.6}$$

where δ_{k-1} is error between measured state & true state.
Batch-mode equation by combining Equ.(4.5) & (4.6),

$$\begin{bmatrix} z_k \\ \hat{x}_{k|k-1} \end{bmatrix} = \begin{bmatrix} H_k \\ I \end{bmatrix} x_k + \begin{bmatrix} e_k \\ \delta_{k-1} \end{bmatrix} \tag{4.7}$$

where I represent the identity matrix.Equation () can be represented as,

$$\tilde{z} = \tilde{H}x_k + \tilde{e}_k \tag{4.8}$$

where, Observation matrix

$$\tilde{z} = \tilde{H}x_k + \tilde{e}_k \tag{4.9}$$

and
error matrix

$$\tilde{e}_k = \begin{bmatrix} e_k \\ \delta_{k-1} \end{bmatrix} \tag{4.10}$$

The covariance matrix $\tilde{R}_k$ of this error matrix is given by

$$\tilde{e}_k = \begin{bmatrix} e_k \\ \delta_{k-1} \end{bmatrix} \tag{4.11}$$

where R_k is noise covariance of $e_k$and error covariance matrix after prediction $\tilde{P}_{k|k-1}$ is represented by,

$$\tilde{P}_{k|k-1} = F_k \tilde{P}_{k-1|k-1} F_k^T + W_k \tag{4.12}$$

But wee must careful that it should satisfy the filter assumption . so decorrelate the data by pre-whitening before solving by state estimation.

different type of pre-whitening methods are available in literature ,Before pre-whitening we must find the exact position of outlier & handle them. Because if outlier will present in the sample then it may cause negative effect. For linear regression model effect of classical pre-whitening in the outlier discuused here,

$$z = Hx + e \tag{4.13}$$

where,

$$H = \begin{bmatrix} h_{11} & . & . & . & h_{1n} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ h_{m1} & . & . & . & h_{mn} \end{bmatrix} = \begin{bmatrix} h_1^T \\ \\ \\ \\ \\ h_m^T \end{bmatrix} \tag{4.14}$$

& each h_i is defined as point of N dimensional vector which follows the Normal distribution $\sim$ N$(\bar{h}, R)$. For this vector unbiased covariance matrix is given as

$$\hat{R} = \frac{1}{m-1} \sum_{i=1}^{m} \left(h_i - \bar{h}\right)\left(h_i - \bar{h}\right)^T \tag{4.15}$$

where sample mean is given by

$$\bar{h} = \frac{1}{m} \sum_{i=1}^{m} h_i \tag{4.16}$$

For ideal decorrelated data covariance matrix is given by,

$$R_{IDEAL} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.17}$$

But data available is not perfectly deccorelated , let us consider a data set of

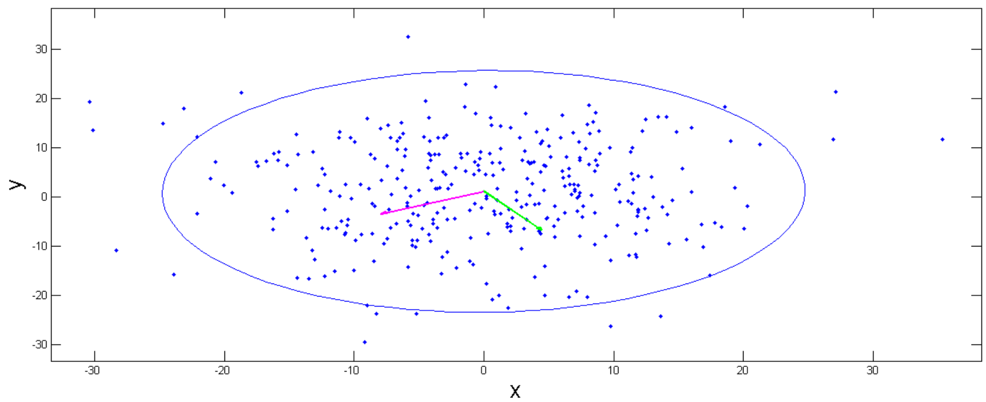


Figure 4.1: For correlated gaussian data without outlier 97.5% confidence ellipse

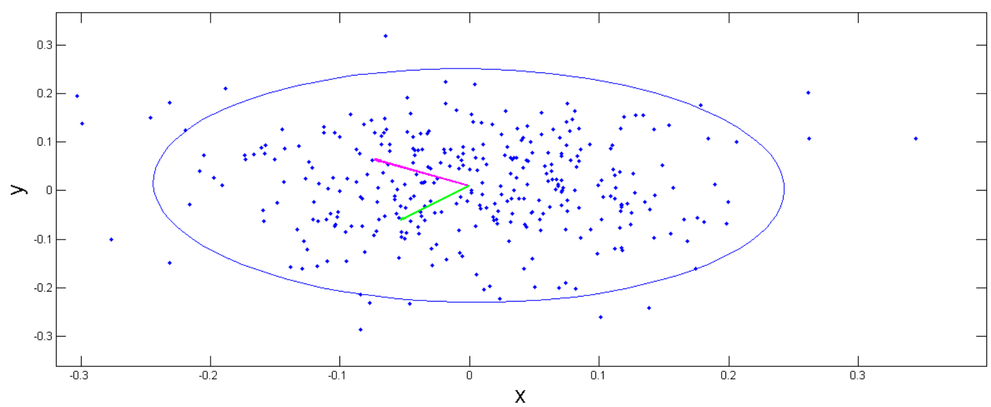


Figure 4.2: For prewhitened gaussian data without outlier 97.5% confidence ellipse

vector m=100 , n=2 .

$$h_1, \ldots\ldots, h_{100} = \begin{bmatrix} 0.218 & .753 & 2.69 & . & . & . & . & -2.13 & -1.02 & 1.7878 \\ -1.407 & -.7399 & .1715 & . & . & . & . & -4.79 & 0.334 & .8797 \end{bmatrix}$$

The associated covariance matrix

$$R = \begin{bmatrix} 102.488 & 1.4876 \\ 1.4876 & 100.730 \end{bmatrix} \tag{4.18}$$

For this 97.5% confidence ellipse is with some correlation is shown in fig. decorrelated data can be found by applying classical prewhitening method in the correlated data. Equation for classical prewhitening is given by:

$$h_{whi} = R^{-1}\left(h_i - \bar{h}\right) \tag{4.19}$$

where h_{whi} is decorrelated data. for this associated covariance matrix is given

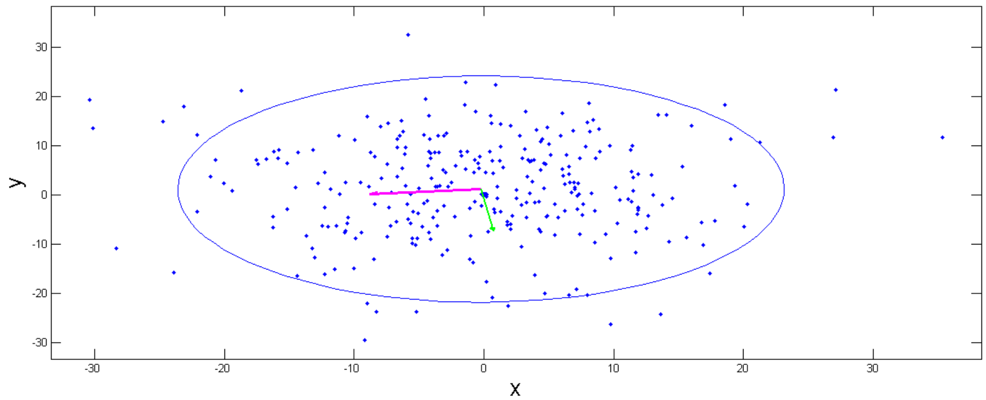


Figure 4.3: For correlated gaussian data with outlier 97.5% confidence ellipse

as

$$R = \begin{bmatrix} .0098 & -.0002 \\ -.0002 & .0096 \end{bmatrix} \tag{4.20}$$

This method performs well if outlier is absent in samples as shown in fig . If we place 15 outlier in same sample (i.e. 5%) . This oulier may be introduce in data from faulty observation or hardware; covariance matrix for above data is given as 97.5% confidence ellipse for data having outlier,is shown in fig . so we do not require this type of erroneous data, with the help of covariance matrix , we can do prewhitening of data & for that covariance matrix is given by:

$$R = \begin{bmatrix} .0113 & -.0001 \\ -.0001 & .0109 \end{bmatrix} \tag{4.21}$$

97.5 % confidence ellipse after prewhitening of outlier presented data is shown in fig . It is clear from fig if outlier present in data then classical whitening might not be suitable to decorrelate the data. For this we need to remove the outlier from sample data.

4.4 Projection Statistic

The projection statistic can be represented as

$$PS_i = \max_{\|u\|=1} \frac{\left| h_i^T u - \underset{j}{med}\left(h_j^T u\right) \right|}{1.4826 \underset{i}{med} \left| h_i^T u - \underset{j}{med}\left(h_j^T u\right) \right|} \tag{4.22}$$

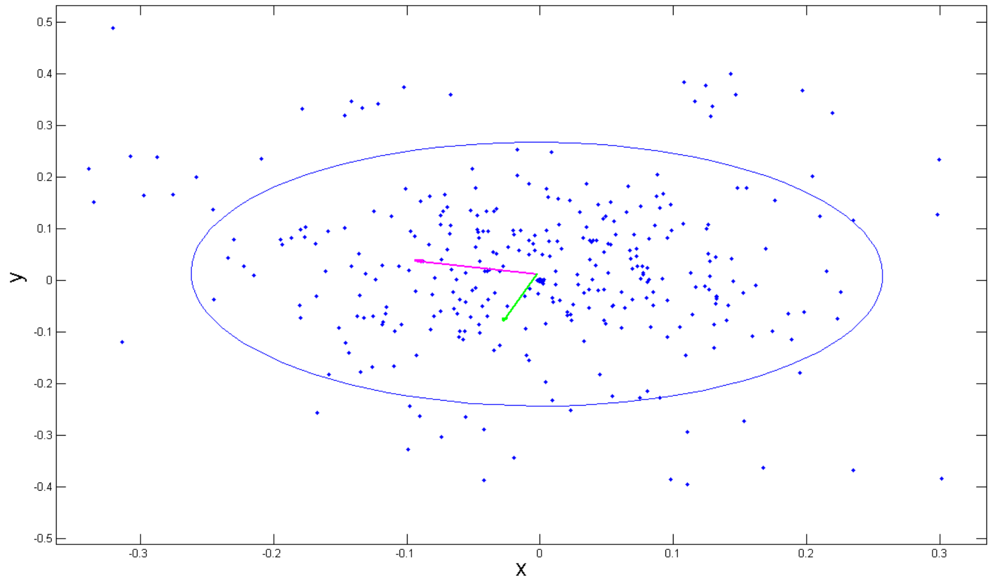


Figure 4.4: for prewhitened gaussian data with outlier97.5% confidence ellipse

It is not practically possible to take projection in every direction . so In this we consider that point only whose vector u generate coordinate wise median and pass through one of data points h_i.

1. These are procedure for the projection statistic algorithm to detect the position of outlier:

   (a) Calculate coordinate-wise median of vector h_j where j=1,2,.....m

   $$m = \left\{med\left(h_{j1}\right), med\left(h_{j1}\right) ...., med\left(h_{jn}\right)\right\} \tag{4.23}$$

   (b) calculate normlized direction

   $$u_j = \frac{h_j - m}{\left\|h_j - m\right\|} \; for \;\; j = 1, 2, ..., m \tag{4.24}$$

   (c) For each direction vector u_j

      (I) In these direction vector project data vectors **h**,represented as

      $$z_{1j} = h_1^T u_j; z_{2j} = h_2^T u_j; ....; z_{mj} = h_m^T u_j; \tag{4.25}$$

      (II) For each direction j find median $z_{med,j} = \left\{z_{1j}, z_{2j}, ..., z_{mj}\right\}$

(d) Median absolute deviation (**MAD**) is given by

$$\mathbf{MAD} = 1.4826 \underset{i}{med} \left| z_{ij} - z_{med,j} \right| \tag{4.26}$$

(e) Standardized projection is given as,

$$P_{ij} = \frac{\left| z_{ij} - z_{med,j} \right|}{MAD} \; for \; i = 1, 2, ..., m \tag{4.27}$$

(f) Repeat the step (3) and calculate standardized projection as:

$$\{P_{i1}, P_{i2}, P_{i3}, ..., P_{im}\} \; for \; i = 1, 2, ..., m \tag{4.28}$$

(g) Finally Calculate the projection statistics as

$$PS_i = \max\{P_{i1}, P_{i2}, P_{i3}, ..., P_{im}\} \; for \; i = 1, 2, ..., m \tag{4.29}$$

2. Calculate the weight function w_i to downweight the outlier, those are found using projection statistics, by

$$w_i = \min\left(1, \frac{d^2}{{PS_i}^2}\right) \tag{4.30}$$

3. Calculate covariance $\tilde{R}_k$ as

$$\tilde{R}_k = \begin{bmatrix} R_k & 0 \\ 0 & \tilde{P}_{k|k-1} \end{bmatrix} \tag{4.31}$$

In KF this can be calculated byW_k & R_k.

4. By cholesky decomposition find C_k such that

$$\tilde{R}_k = C_k C_k^T$$

This can also be done by square root method,given as

$$\tilde{R}_k = \sqrt{\tilde{R}_k}\sqrt{\tilde{R}_k}$$

5. Multiply the linear regression model $\tilde{z} = \tilde{H}x_k + \tilde{e}_k$ by ${C_k}^{-1}$ or $\left(\sqrt{\tilde{R}_k}\right)^{-1}$ given as,

$$\left(C_k^{-1}\right)\tilde{z}_k = \left(C_k^{-1}\right)\tilde{H}_k x_k + \left(C_k^{-1}\right)\tilde{e}_k \tag{4.32}$$

6. After solving Equ. it provide the equation in following form

$$y_k = A_k x_k + \eta_k \tag{4.33}$$

4.5 Robust Filtering based on GM-estimation

To solve the linear regression equation

$$y_k = A_k x_k + \eta_k$$

An M-estimator minimizes the objective function given below

$$\xi(x) = \sum_{i=1}^{m} \rho\left(\frac{r_i}{s}\right) \tag{4.34}$$

where $\rho(.)$ is a nonlinear function, s is robust scale estimate while r_i is residual. One popular class of M-estimator is Huber estimator given by

$$\rho\left(\frac{r_i}{s}\right) = \begin{cases} \frac{1}{2}\left|\frac{r_i}{s}\right|^2 \ for \ \left|\frac{r_i}{s}\right| < \delta \\ \delta\left|\frac{r_i}{s}\right| - \frac{\delta^2}{2}, \ elsewhere \end{cases} \tag{4.35}$$

where $\delta = 1.5$ for good statistical efficiency. The residual vector for i^{th} element is represented by

$$r_i = y_i - a_i^T \hat{x} \tag{4.36}$$

Robust scale estimator s is defined as

$$s = MAD = 1.4826 median_i |r_i| \tag{4.37}$$

Taking partial derivative & equating with zero we get the solution of objective function as

$$\frac{\partial \xi(x)}{\partial x} = \sum_{i=1}^{m} \frac{1}{s} \frac{\partial \rho\left(\frac{r_i}{s}\right)}{\partial\left(\frac{r_i}{s}\right)} \left(\frac{\partial r_i}{\partial x}\right)^T = 0 \tag{4.38}$$

$$= \sum_{i=1}^{m} -\frac{a_i}{s} \frac{\partial \rho\left(\frac{r_i}{s}\right)}{\partial\left(\frac{r_i}{s}\right)} = 0 \tag{4.39}$$

score fuction $\phi\left(\frac{r_i}{s}\right)$ is used in above equation and by solving the equation with differentiability property we get

$$\sum_{i=1}^{m} \frac{a_i}{s} \phi\left(\frac{r_i}{s}\right) = 0 \tag{4.40}$$

For huber function, score function $\phi\left(\frac{r_i}{s}\right)$ is given by

$$\phi\left(\frac{r_i}{s}\right) = \begin{cases} \left(\frac{r_i}{s}\right), \quad for \ \left|\frac{r_i}{s}\right| < \delta \\ \delta \times sign\left(\frac{r_i}{s}\right), \ elsewhere \end{cases} \tag{4.41}$$

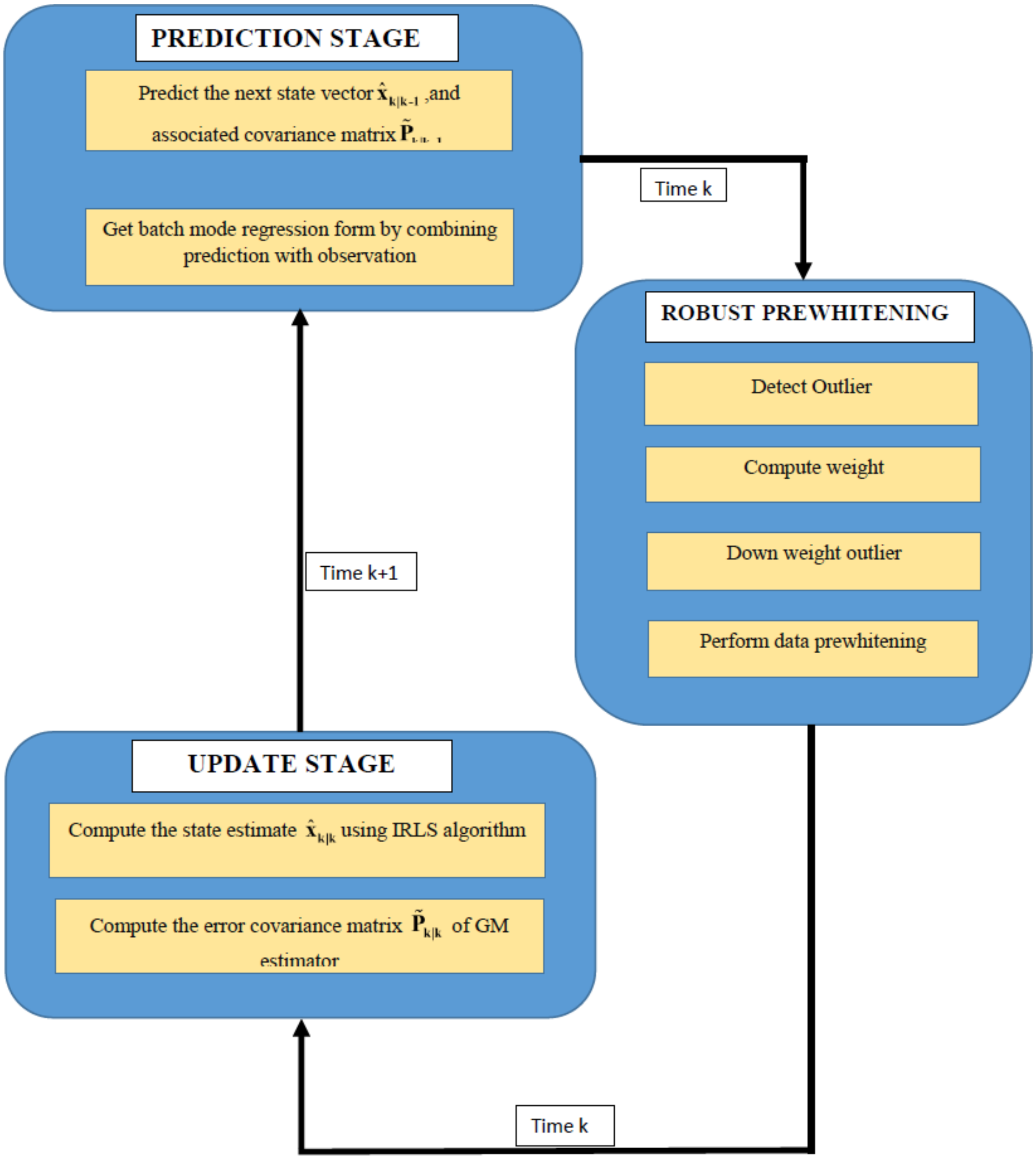


Figure 4.5: Flow chart of working of GM-KF Algorithm

Solve the equation () by iteratively Re-weighted Least Squares (IRLS) algorithm as follows:

$$\sum_{i=1}^{m}\left(\frac{a_i}{s}\right)\left(\frac{y_i - a_i^T\hat{x}}{s}\right)q\left(\frac{r_i}{s}\right) = 0 \tag{4.42}$$

where

$$q\left(\frac{r_i}{s}\right) = \frac{\phi\left(\frac{r_i}{s}\right)}{\left(\frac{r_i}{s}\right)} \tag{4.43}$$

Writing above equation in matrix form, we get

$$A^T Q(y - A\hat{x}) = 0 \tag{4.44}$$

$$A^T Qy - A^T QA\hat{x} = 0 \tag{4.45}$$

where

$$Q = diag\left\{q\left(\frac{r_i}{s}\right)\right\} \tag{4.46}$$

Final of M-kalman filter for linear regression equation is given by

$$\hat{x}_{k|k}^{l+1} = \left(A^T Q^{(l)} A\right)^{-1} A^T Q^{(l)} y_k \tag{4.47}$$

Chapter 5

Simulation Results

Vehicle tracking Controller

Aircraft Tracking Model tracking Controller

Chapter 5
Simulation Results

5.1 Vehicle tracking Controller

The main advantage of GM-KF is its ability to reject outliers. First consider a model for GPS-based vehicle controller with transition and observation matrices:

$$F = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.1}$$

$$\begin{bmatrix} & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ H = & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}^T \tag{5.2}$$

Assumption for given model are as follows; sampling rate=5Hz that gives sampling period of 0.2 sec. Gaussian noise with observation covariance matrix **R** and system noise covarince **W** is given by

$$R = \begin{bmatrix} .01 & 0 & 0 & 0 \\ 0 & .01 & 0 & 0 \\ 0 & 0 & .01 & 0 \\ 0 & 0 & 0 & .01 \end{bmatrix} \quad and \quad W = \begin{bmatrix} .01 & 0 & 0 & 0 \\ 0 & .02 & 0 & 0 \\ 0 & 0 & .03 & 0 \\ 0 & 0 & 0 & .04 \end{bmatrix}$$

from above it is clear that m=8 and n=4.To check whether this filter is robust or not place one outliers with value of -40 at second element of z at t=40sec,when position of vehicle is x1=84 and x2=-40.As it can observe in the figure that GM-

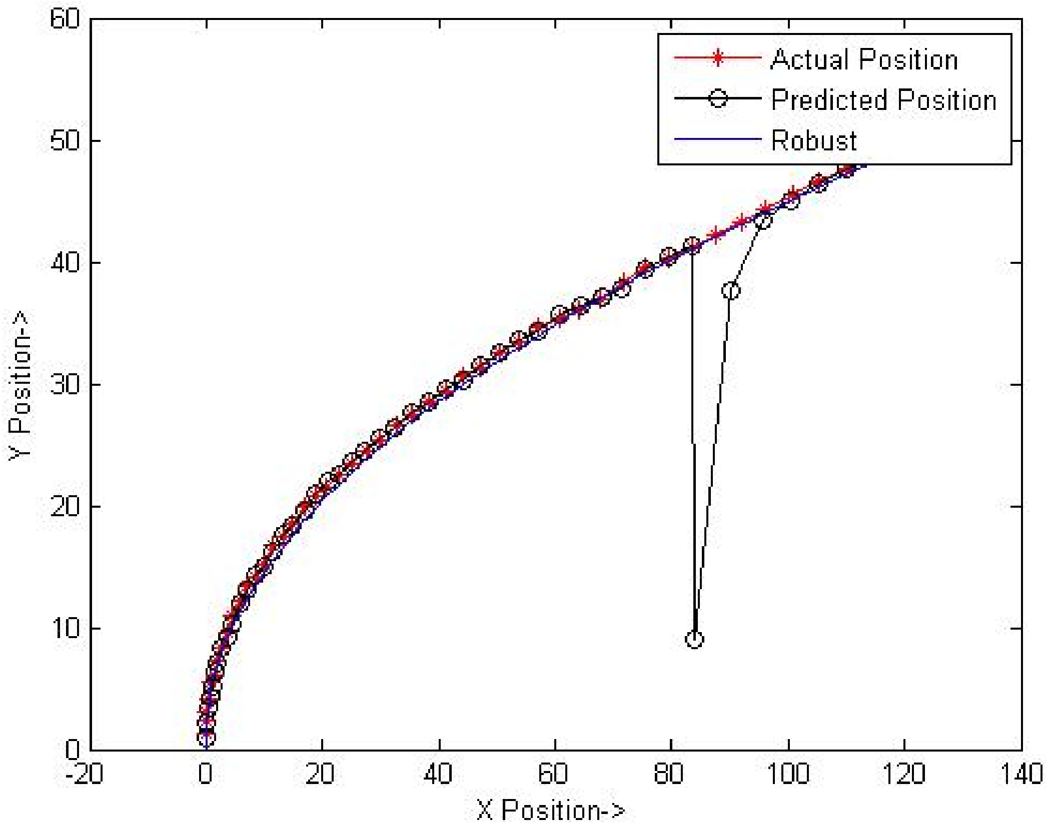


Figure 5.1: GPS based vehicle tracking in the presence of outlier

KF withstand in the presence of observation outliers. the weight matrix Q of the GM-estimation is used to increase the performance of estimator.By combining observation and multiple prediction measurement we can define robust weight superior $\varepsilon^* = \left(\frac{m}{n} >> 1\right)$

5.2 Aircraft Tracking Model

For an aircraft in a circular maneuvering exercise consider a tracking problem using GPS data.The state transition matrix **F** for the given model is

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0.1 \\ 0 & 0 & -0.9 & 1 \end{bmatrix} \tag{5.3}$$

and observation matrix **H** for this model is defined as

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}^T \tag{5.4}$$

For this model, the maximum number of outliers that any equivalent estimator

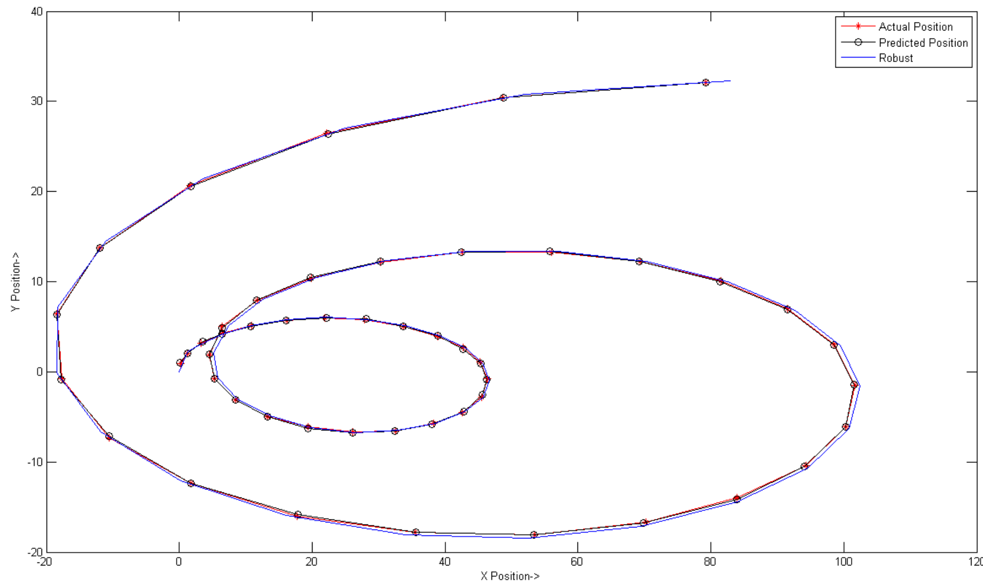


Figure 5.2: performance comparision of KF & GM-KF if no outlier present

can handle is [(m + n) n)/2] = [m/2] = 8/2 = 4, yielding a maximum breakdown point of ε^*= 4/12 = 0.33.Through different simulation by adding innovation , observation and structural outlier for a given time period introduced simultaneously ans for several time step successively. Through this simulation we found that GM-KF can suppress 3 simultaneous outlier but not 4 which gives breakdown point of ε^* =3/12 = 0.25, that is little less than maximum breakdown point. Now consider a case where observation covariance matrix R & system noise covariance W is given by,

$$R = \begin{bmatrix} .01 & 0 & 0 & 0 \\ 0 & .01 & 0 & 0 \\ 0 & 0 & .01 & 0 \\ 0 & 0 & 0 & .01 \end{bmatrix} \quad and \quad W = \begin{bmatrix} .01 & 0 & 0 & 0 \\ 0 & .01 & 0 & 0 \\ 0 & 0 & .01 & 0 \\ 0 & 0 & 0 & .01 \end{bmatrix}$$

In this model total 3 outlier occurs at t=20s to t=23s i.e. one innovation outlier at first element of predicted vector

$$\left[\hat{x}^1_{20|21}, \hat{x}^1_{21|22}, \hat{x}^1_{21|22}\right] = [-32.6, -38.3, -45.1]$$

is replaced by

$$\left[\hat{x}^1_{20|21}, \hat{x}^1_{21|22}, \hat{x}^1_{21|22}\right] = [-402.6, -5.3, -1.1]$$

and two observation

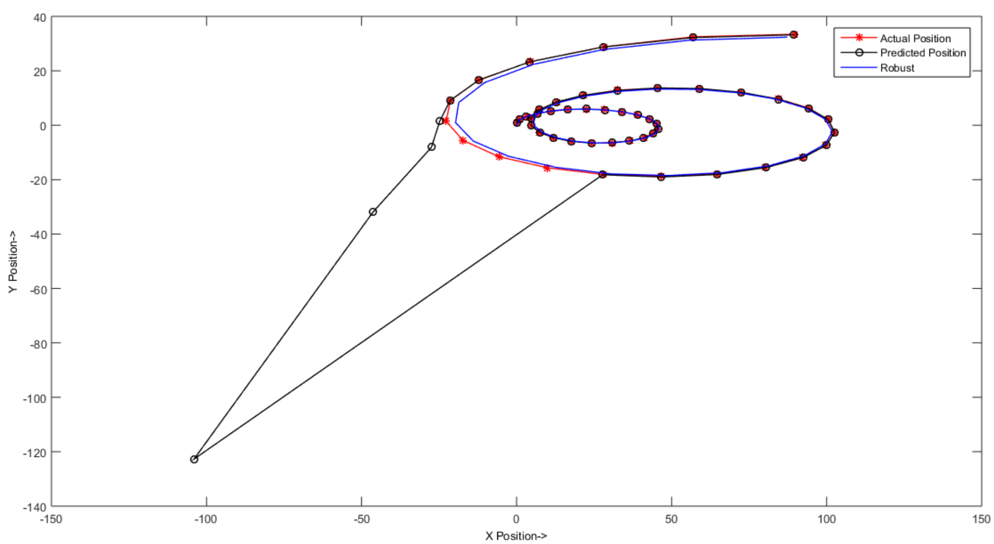


Figure 5.3: performance comparison of KF & GM-KF if outlier present

$$\begin{bmatrix} z^1_{20}, z^1_{21}, z^1_{22}, z^1_{23} \\ z^2_{20}, z^2_{21}, z^2_{22}, z^2_{23} \end{bmatrix} = \begin{bmatrix} -5.7, 8.3, -12.3, -17.8 \\ -32.1, -37.9, -42.1, -49.3 \end{bmatrix}$$

are replace by

$$\begin{bmatrix} z^1_{20}, z^1_{21}, z^1_{22}, z^1_{23} \\ z^2_{20}, z^2_{21}, z^2_{22}, z^2_{23} \end{bmatrix} = \begin{bmatrix} -11.7, -18.3, -22.1, -27.8 \\ -41.1, -26.9, -22.1, -62.3 \end{bmatrix}$$

From figure(5.3) it is clear that GM-KF can suppress all 3 types of outlier, so it can provide true state even in the presence of outliers.

Chapter 6

Development & application of GM-EKF

System with multiple equilibrium point

Extended Kalman Filter

The Langevin Model

Appliaction of Extended kalman filter in Langevin Model

Chapter 6

Devlopment & application of GM-EKF

If noise is Gaussian then classical kalman filter is optimal estimator but for non Gaussian we develop a new type of estimator,GM-KF which is robust to different type of outlier . Now this filter can also be used for Non-linear system with additive Gaussian noise known as Extended Kalman Filter. In this chapter a new GM-EKF is proposed which is robust to different type outliers in nonlinear system with multiple equilibrium point[8].

6.1 System with multiple equilibrium point

Nonlinear system having multiple equilibrium point studied in this chapter. This type systems have small fluctuations for some time around one equilibrium point and it will shift suddenly to another equilibrium point.In solid mechanics example of this type of system is Buckling of elastic beam . If there is no outside pressure then system will rest in its original position[17, 16]. If any outside compression force will be applied then the beam will deform according to Hooke's, this effect can be seen as fluctuations . If the compression force is large then the linear theory fails and buckling effect occurs . Because of this fluctuation around equilibrium point will increase and cause system to shift from one equilibrium point to another.

Again for this model we assume noise with Gaussian PDF , But if noise is Non-Gaussian including innovation & observation outliers. This type of noise may lead to faulty estimation, so we need a robust estimator that perform well in the presence of outliers and should be able to provide reliable estimate over time. The KF is designed for linear system so it is not able to track effectively in non linear sytem and also not robust to outliers[19].

6.2 Extended Kalman Filter

for state estimation in nonlinear systems one popular method is the extended Kalman filter. EKF assume non linear model in the prediction stage. Linearization and discretization of this nonlinear model are performed around previous stage to get discrete, linear equations. from the classical KF recursion we can derive EKF equation . It gives good result under the assumption while in the presence of outliers its performance is poor[6, 8].

This poor performance of EKF is because of

- It develop under the weak noise assumption, it does not track system satisfactory in the presence of the Gaussian noise, even it diverge in the presence of outliers.
- It strongly depend on accurate observation to track transition causes by system process noise.
- It strongly depend on observation so it may undergo the transition by ignoring correct prediction.
- If system process noise is comparatively smaller than the observation noise, then EKF miss a system transition by ignoring good observation while incorrectly relying in prediction.

To overcome all weak performance cause of EKF we use a new method that reliably track the state of nonlinear dynamic system.To design EKF we follow 3 step similar to GM-KF which include (I) By creating a redundant observation vector,(II) robust prewhitening of observatoin,(III) estimating the state vector robustly.consider an example to track climate transition quickly and reliably i.e. Langevin Model[8].

6.2.1 Review of Extended KF

The state equation of system dynamic is given by

$$\dot{x}_t = f(x_t) + w_t + u_t \tag{6.1}$$

$$y_t = h(x_t) + e_t \tag{6.2}$$

where $f(x_t)$ & $h(x_t)$ is continuously differentiable with respect to x_t. Noise co-variance is denoted by W_t & R_t . For nonlinear equation Taylor series expansion is given by

$$f(x_t) = f(x_t^*) + A_x(x_t^*)\,\delta x_t \tag{6.3}$$

where A_x is Jacobian matrix given as,

$$A_x = \left.\frac{\partial f(x_t)}{\partial x_t}\right|_{x_t = x_t^*} \tag{6.4}$$

δx_t is given by

$$\delta x_t = x_t - x_t^* \tag{6.5}$$

from equation (6.1) & (6.3) we will get state equation in form of

$$\delta \dot{x}_t = A_x(x_t^*)\,\delta x_t + w_t + u_t \tag{6.6}$$

similarly equation (6.2) represented as

$$\delta z_t = H_x(x_t^*)\,\delta x_t + e_t \tag{6.7}$$

where H_x is Jacobian matrix given as,

$$H_x = \left.\frac{\partial h(x_t)}{\partial x_t}\right|_{x_t = \dot{x}_t^*} \tag{6.8}$$

This nonlinear model cab be discretized and represented as

$$\begin{aligned} x_{k+1} &= A_d x_k + w_k + B_d u_k \\ y_k &= H_d x_k + e_k \end{aligned} \tag{6.9}$$

where

$$A_d = e^{A_x T_s}$$

$$H_d = H_x$$

For this recursion can be written as

$$\hat{x}_{k|k-1} = A_d \hat{x}_{k-1|k-1} + \int_{t_{k-1}}^{t_k} f\left(\hat{x}_{k-1|k-1}\right) dt + B_d u_k \tag{6.10}$$

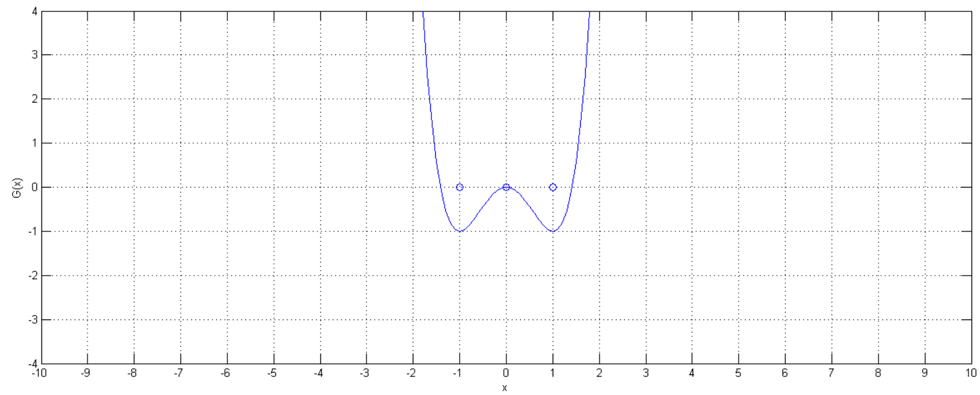


Figure 6.1: 3 equilibrium point of Potential function G(x)

$$\tilde{P}_{k|k-1} = A_d \tilde{P}_{k-1|k-1} A_d^T + W_k \tag{6.11}$$

$$K_k = \tilde{P}_{k|k-1} H_d^T \left[H_d \tilde{P}_{k|k-1} H_d^T + R_k \right]^{-1} \tag{6.12}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left[y_k - h\left(\hat{x}_{k|k-1}\right) \right] \tag{6.13}$$

$$\tilde{P}_{k|k} = \tilde{P}_{k|k-1} - K_k H_d \tilde{P}_{k|k-1} \tag{6.14}$$

6.3 The Langevin Model

A nonlinear system with multiple equilibrium point is represented by equation

$$\dot{x}_t = f(x_t) = -4x_t(x_t^2 - 1), \tag{6.15}$$

The negative gradient of above equation is expressed as,

$$G(x_t) = x_t^4 - 2x_t^2 \tag{6.16}$$

the dynamic equation can be written as in presence of noise,

$$\dot{x}_t = 4x_t - 4x_t^3 + \eta_t \tag{6.17}$$

The potential function given by (6.16) is shown in figure (6.1) while the equation (6.15) of f(x) is represented in figure (6.2). If there is no external force or noise then this system has three equilibrium point at x=0,x=1 & x=-1. The first one is unstable while the later two is stable equilibrium point[17]. Under normal condition it will fluctuate around one of this equilibrium point if no external force applied. If any external force with large enough magnitude is applied then it will shift from one equilibrium point to another, called transition

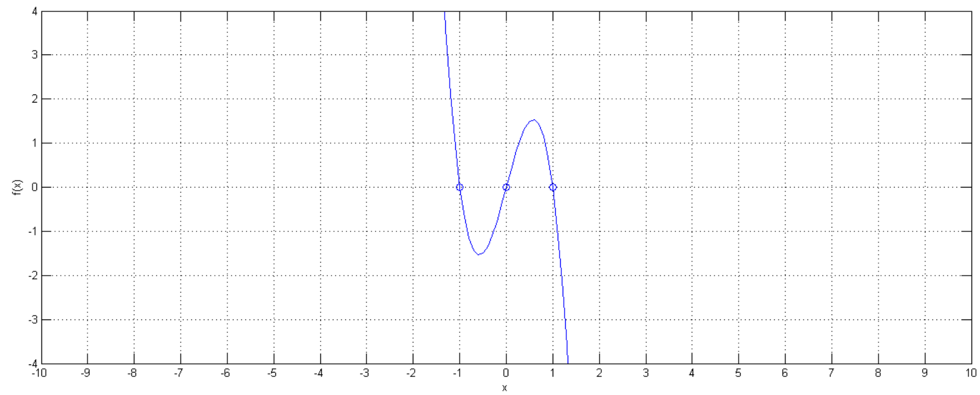


Figure 6.2: 3 equilibrium point of system dynamic equation $f(x) = -G'(x)$

in the system.

The function $G(x_t)$ given in equation (6.16) will consider as climate potential where x_t represent the erth's saverage surface temperature at time t. In the figure one equilibrium point represent climate state and the otherone represent ice ages[16, 6].

6.4 Appliaction of Extended kalman filter in Langevin Model

In langevin equation we now apply EkF and evaluate the performance how well it tracks the transition.Under some condition it performs well and fails to track in some condition.

Case 1: Statistical Efficiency

we evaluate the performance of filter under Gaussian distribution and find how well it estimate under normal assumption.As expected the filter perform well under the Gaussian noise distribution which can be shown in bottom of figure with low MSE[6].The noise assumption made for this case is observation noise $\sigma_y^2 = 0.01$ and system process noise $\sigma_x^2 = 0.01$

Case 2: Effect of Noise variance in system transition

Here we see that how well EKF track transition of system from one equilibrium point to another. For this simulation observation noise $\sigma_y^2 = 0.01$ and system process noise $\sigma_x^2 = 0.001$ sampling period =0.25sec and cor-

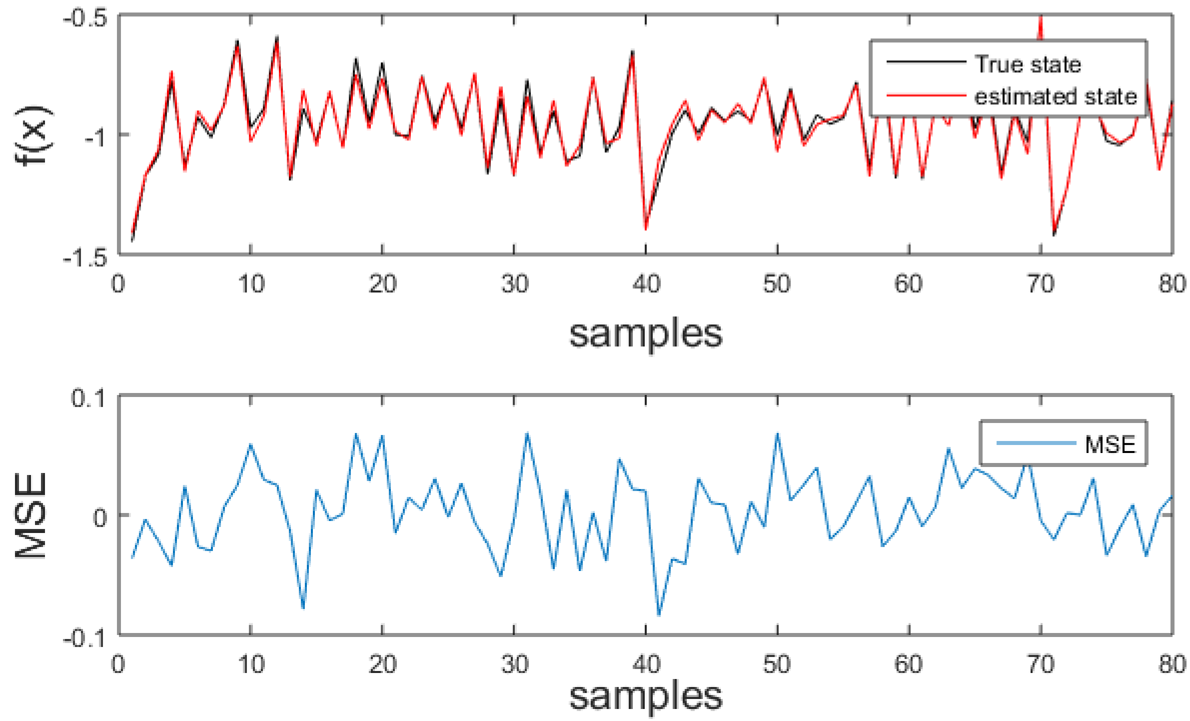


Figure 6.3: Case 1 - For double well system performance of EKF.

responding frequency 4 Hz. fig6.5 shows that for new prediction step sample available at 0.25sec with corresponding sampling frequency of 4 hz. It tracks the transition with delay of 1 sec.If the observations are accurate enough to make this filter gain greater than 0.50, then the filter is expected to correctly track the system transition[6].

If observation noise is sufficiently larger then system process noise, then it may cause the filter gain just not large enough for a long period and causing the filter too confident about the prediction and provide inaccurate state . Because of this filter can not track the state after 160 observation i.e. 40 sec as shown in figure(6.6).

case 3: Transition from one equilibrium point to another

in figure (6.7) it has shown that transition of state from equilibrium point +1 to -1 and back again within 6 samples. if $\sigma_y^2 = 0.04$ & $\sigma_x^2 = 0.01$ then it can track the transition after 5 sample i.e. after delay of 20 sec. If

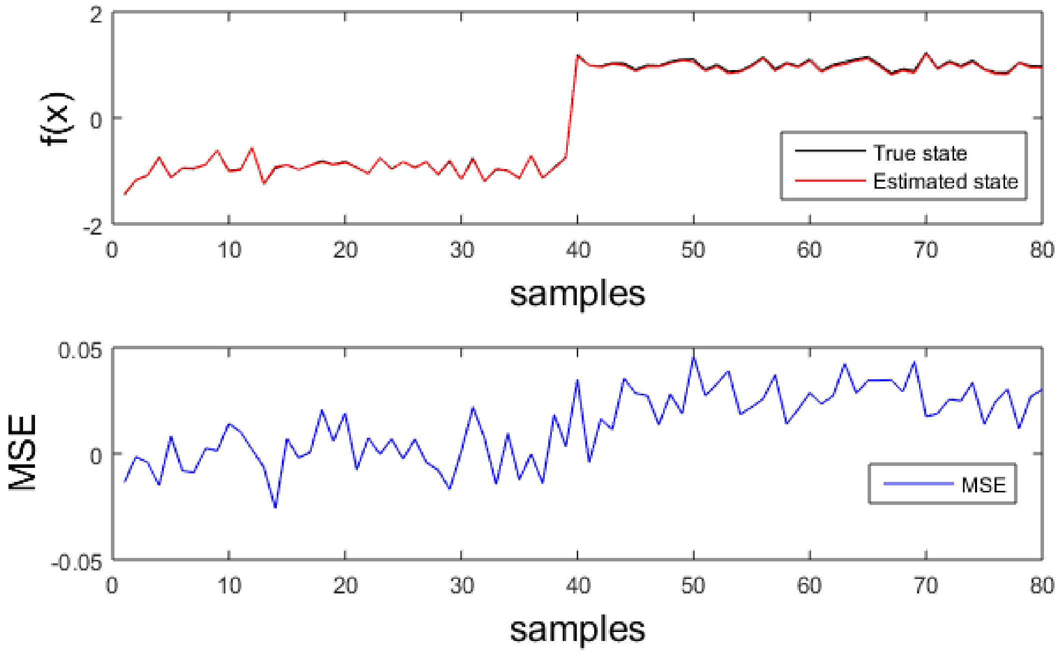


Figure 6.4: Case 2 - The EKF tracks state estimation transition with a delay of 1 seconds for observation frequency 4Hz

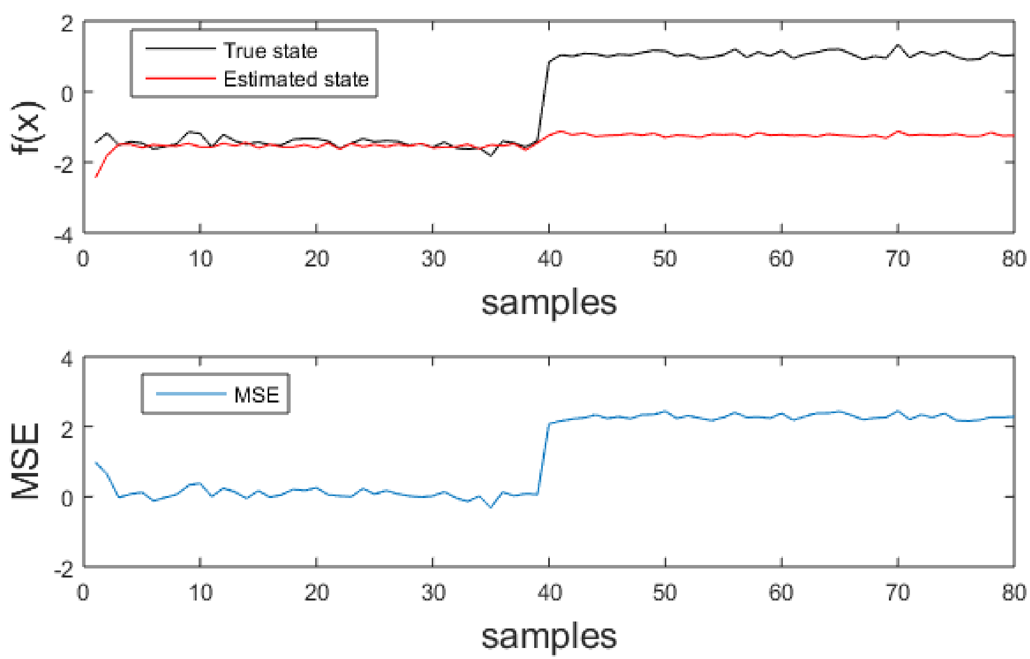


Figure 6.5: The EKF not able to track transition even after 40 seconds when $\sigma_y^2 = 0.08$ and $\sigma_x^2 = 0.01$

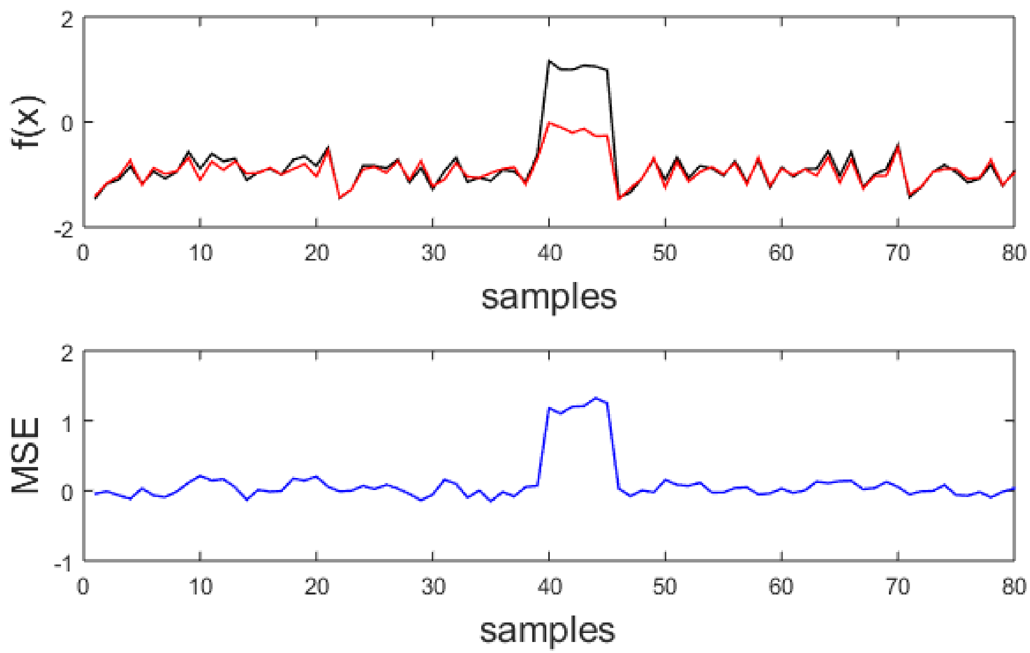


Figure 6.6: Case 4 - IF observation noise is sufficiently low then performance of EKF, i.e. $\sigma_y^2 = 0.03$ and $\sigma_x^2 = 0.01$ it with a noticeable delay

$\sigma_y^2 = 0.06$ & $\sigma_x^2 = 0.01$ the EKF does not track the transition, as shown in Figure 6.8 the latter does not last long enough either to process a sequence of observations that may eventually force the filter to switch to another equilibrium point. Therefore, the EKF completely fails to follow the system state.

case 4: Performance in the presence of outliers

Now let us consider the case where one or more than one observation containing outliers[10], in this result outliers is from t=39 sec to t=41 sec.In this case EFF is unable to distinguish whether observation is good or bad and gives inaccurate state estimate as shown in figure (6.9) .

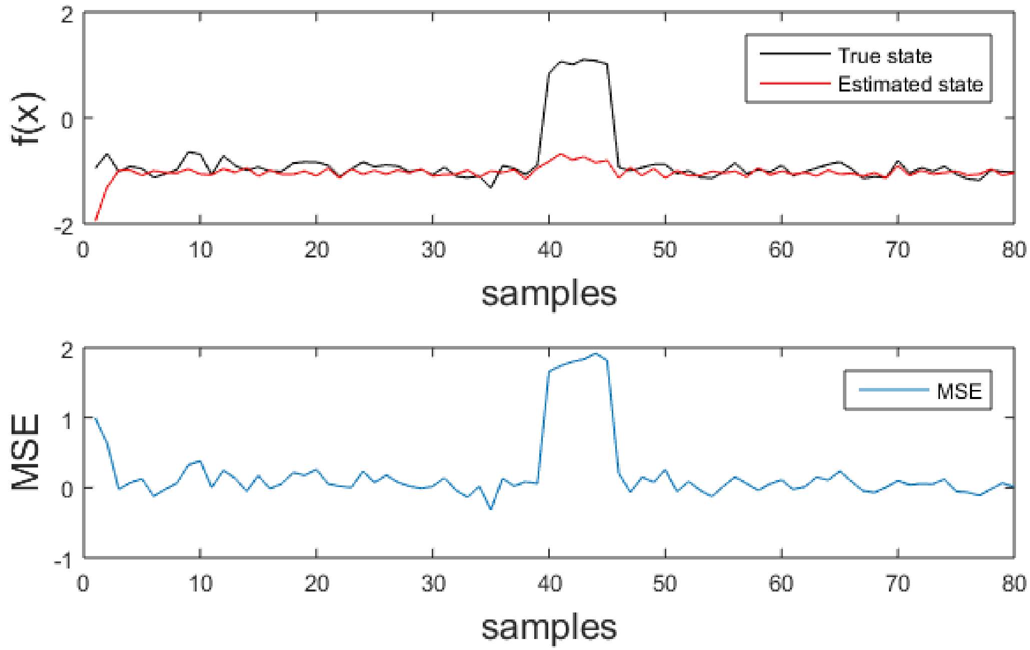


Figure 6.7: Case 3 - EKF unable to track transition after certain value of observation noise, in this case $\sigma_y^2 = 0.06$ with $\sigma_x^2 = 0.01$

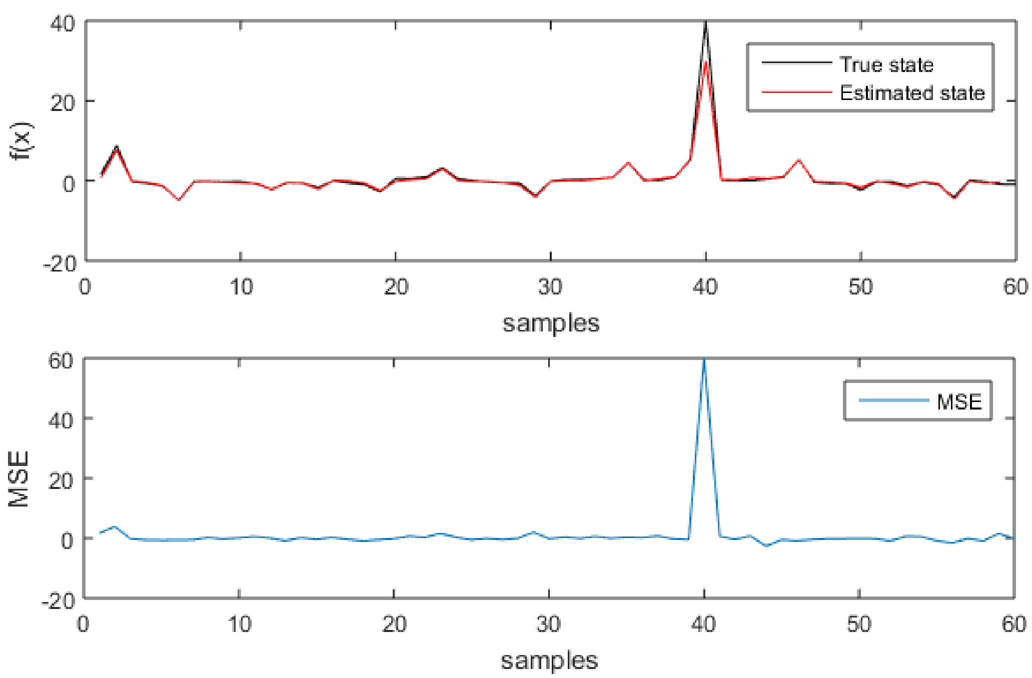


Figure 6.8: Case 4 - In the presence of outliers EKF gives inaccurately estimation

So from this it is clear that EKF perform inaccurately in the presence of outliers so we need a Robust filter to overcome such limitation with some advantage. Firstly, to overcome observation sensitivity we need redundant observation . It also overcomes the overconfidence in prediction. Second, to make filter robust we need robust estimate of variance and weight with reliable weighting algorithm. We can use PS estimator that we used in case of GM-KF[8].

Chapter 7
Conclusion & Future work

In this research we develop a new type of filter for system model component such as noise, input, transition matrix for state,observation matrix and which is affected by structural, innovation and observation outlier.A generalized maximum likelihood type estimator is developed to achieve robustness in various aspect by new data prewhitening method and redundancy in observation vector.

From Robust statistical theory we adopt concept of redundancy and breakdown point. Redundancy can be achieved by taking batch mode observation for linear regression by combining state prediction and observations. To handle the different type of outlier and decorrelating the data a new prewhitening concept is used.Linear regression allos us to use any type of estimator for Robust estimation whose covariance matrix can be derived. To solve state estimation equation IRLS algorithm and GM-estimator has been used. To suppress the different kind of outlier such as vertical outlier and bad leverage point we must characterize this outlier.

Several other research topics on the GM-KF and GM-EKF can be formulated. we may apply the stabilization of KF and EKF to the new filter. WE may find application of this filter if outlier will present in the other part of system such as in covariance matrices R and W. We can use another outlier detection method in place projection statistics.In nonlinear regression we can study the effect of outlier because in this definition of breakdown point for linear system is not applied. Furthe in place of IRLS algorithm to solve state equation we can use newton iteration method.

Bibliography

[1] Vic Barnett and Toby Lewis. *Outliers in statistical data*, volume 3. Wiley New York, 1994.

[2] Richard S Bucy and Kenneth D Senne. Realization of optimum discrete-time nonlinear estimators. Technical report, DTIC Document, 1971.

[3] James V Candy. *Model-based signal processing*, volume 36. John Wiley & Sons, 2005.

[4] Chung Chen and Lon-Mu Liu. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421):284–297, 1993.

[5] Zeljko M Durovic and Branko D Kovacevic. Robust estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, 44(6):1292–1296, 1999.

[6] GREGORY L Eyink and Juan M Restrepo. Most probable histories for nonlinear dynamics: tracking climate transitions. *Journal of Statistical Physics*, 101(1-2):459–472, 2000.

[7] Ronald Aylmer Fisher. Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 22, pages 700–725. Cambridge Univ Press, 1925.

[8] Mital A Gandhi and Lamine Mili. Robust extended kalman filter for transient tracking and outlier suppression.

[9] Mital A Gandhi and Lamine Mili. Robust kalman filter based on a generalized maximum-likelihood-type estimator. *Signal Processing, IEEE Transactions on*, 58(5):2509–2520, 2010.

[10] Peter J Huber. *Robust statistics*. Springer, 2011.

[11] Peter J Huber et al. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[12] Thomas Kailath, Ali H Sayed, and Babak Hassibi. *Linear estimation*, volume 1. Prentice Hall Upper Saddle River, NJ, 2000.

[13] Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. *Journal of Fluids Engineering*, 83(1):95–108, 1961.

[14] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[15] RARD Maronna, Douglas Martin, and Victor Yohai. *Robust statistics*. John Wiley & Sons, Chichester. ISBN, 2006.

[16] Robert N Miller, Everett F Carter, and Sally T Blue. Data assimilation into nonlinear stochastic models. *Tellus A*, 51(2):167–194, 1999.

[17] Robert N Miller, Michael Ghil, and Francois Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the atmospheric sciences*, 51(8):1037–1056, 1994.

[18] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.

[19] Abdelhak M Zoubir, Visa Koivunen, Yacine Chakhchoukh, and Michael Muma. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *Signal Processing Magazine, IEEE*, 29(4):61–80, 2012.