THE TOBIT KALMAN FILTER:
AN ESTIMATOR FOR CENSORED DATA

by

Bethany Allik

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering

Summer 2014

THE TOBIT KALMAN FILTER:
AN ESTIMATOR FOR CENSORED DATA

by

Bethany Allik

Approved: ______________________________
Kenneth E. Barner, Ph.D.
Chair of the Department of Electrical and Computer Engineering

Approved: ______________________________
Babatunde Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: ______________________________
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Ryan Zurakowski, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Abhyudai Singh, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Gonzalo Arce, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Mark Ilg, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

I'd like to thank the Army Research Laboratory for funding this research and providing me with summer and winter internship opportunities. Interning while in graduate school gave me the opportunity to apply my research, and gave me the inspiration to pursue basic research. I'd especially like to thank my mentor and thesis adviser at ARL, Dr. Mark Ilg.

My deepest gratitude is to my adviser at the University of Delaware; Dr. Ryan Zurakowski. His vision provided the framework for the Tobit Kalman filter, while the encouragement and direction he provided to me perfected the results.

I am grateful to Dr. Mike Piovoso, who's advice and insightful comments through long weekly discussions improved the quality of my research immensely.

I'd also like to thank the rest of my thesis committee; Dr. Abhyudai Singh and Dr. Gonzalo Arce. Who both gave practical comments and insights in the proposal and final defense.

Id also like to thank my coworkers at the Army Research Laboratory and the University of Delaware for the support I was given in pursuing my doctorate degree.

A special thanks to my parents; Roberta Dixon and Toomas Allik, who where there for me throughout my college career. Thank you to my brother, Brian Allik, for spending time with me outside of school and obtaining an advanced degree along with me.

Finally, thank you to my best friend and collaborator, Cory Miller.

TABLE OF CONTENTS

Appendix

LIST OF FIGURES

ABSTRACT

The Kalman Filter has become ubiquitous in tracking and estimation. Many estimation applications, especially those using low cost commercial of-the-shelf sensors (COTS), are subject to a special type of measurement nonlinearity called censoring. Censoring frequently takes the form of sensor saturation, occlusion regions, and limit-of-detection. These forms of censoring are known as Tobit model Type 1 censoring. Introduction of censored measurements into the Kalman filter results in biased estimates of the underlying states. In this dissertation, we present the first formulation of the Kalman filter capable of estimating state variables from censored data without bias. We refer to this formulation as the Tobit Kalman filter.

Previous work on Kalman filtering with measurement nonlinearities or sensor faults includes a Kalman filter for intermittent measurements, the particle filter, the unscented Kalman filter (UKF) and the extended Kalman filter (EKF). Intermittent measurement nonlinearity is similar to the censored measurement model; with the exception that censored data measurements are correlated with the state values. Previous work for intermittent measurements in estimation reduces the Kalman filter to a linear predictor when the measurement is missing. Use of either this formulation or a standard Kalman filter as an estimator in a censored data example will result in a biased estimate of the state. The particle filter is able to estimate the state values when the measurements are subject to censoring under certain cases, but comes with a substantial computational burden. The UKF is a less computationally expensive approach that proves to be non-robust when the measurements are near a censoring region. The EKF suffers from an undefined Jacobian at the threshold itself, and the Jacobian is zero in the censored region. On the other hand, the Tobit Kalman filter provides unbiased recursive estimates of latent state variables in or near saturated regions. This results

by properly accounting for the statistics in the Tobit model, and using them to adapt the Kalman filter error terms and state measurement updates. The Tobit Kalman filter is completely recursive and computationally inexpensive while previous attempts at Tobit based state estimators where not recursive and are not computationally feasible.

The many applications to the Tobit Kalman filter include MEMS sensor based tracking with saturation, visual tracking with camera frame censoring and biological measurements with limit of detection saturations. In addition to the theoretical work for the Tobit Kalman filter, this dissertation discusses the applications of the Tobit Kalman filter by presenting simulations.

Chapter 1

INTRODUCTION

In 1958, James Tobin discovered that the statistical relationship between household income and expediture is unique in that the data seemed to fit a linear model mostly, but suffered from clustering of data points at a limiting value [1]. This data clustered at the limiting value, the expediture representing the dependent variable, is refered to as the *censoring* limit on the data. The data clustered at the censoring limit has a unique property in that it does not represent a unique independent variable through a linear operation, but otherwise represents the information that the dependent variable lies within a certain region, the censored region. If censored data is ignored or used in fitting the data to a linear model then the resulting model would be biased.

Tobin's household expenditure problem would typically start with developing a model of this system, however, the existence of several points where certain people would have income but no household expenditure cause a region specific nonlinearity. To fix this, Tobin created the Tobit model [2], named from the fused titles of Tobin and Probit models. This model introduced the idea of censoring as a piecewise nonlinearity in the system model. Since then, the model has been used in other application of economics including income vs inheritance [3]; wages of husbands and wives vs annual days worked [4], technology perception vs adoption [5] and several more [6].

Censoring has been reinvented in biology in the field survival analysis [7] [8]. Survival analysis is a statistical approach used to follow patients over a long period. More specifically, it is a way to model time to an event in biology and has also been used as a measure of time to failure in reliability engineering [9] [10]. Survival analysis has become an important tool in public health monitoring [11]. The basis of survival analysis is the survival function, which is stated as $S(t) = P(t < T)$, where P denotes probability and T denotes time of death, failure etc. Censoring comes into the survival model when the time of death is unknown but known to be after some date; this is referred to as right censoring. This censoring occurs in long period experiments where the patient will miss scheduled check up exams, or clinical trial follow ups [12]. An analogous version of right censoring is left censoring; where the data contains a lower limit. An example of left censoring is when public health data is not observed in children before they have reached kindergarten. In this case where the age limit is known, the ages are censored. However, if the age limits or thresholds of the uncensored regions are unknown, we refer to this as truncation. An example of left truncation would be if young patients are not observed until a certain age, or if some patients had passed before that age then they will not be observed [13].

The censored data model developed independently in economics and biology. For the remainder of this dissertation, the censored data model and terminology used will be based on the Tobit model. The many Tobit models have application in engineering, and are discussed briefly in Section 1.1. The most simplistic version of the Tobit Model contains a latent variable y_t^*, which linearly depends on βx_t,

$$y_t = \begin{cases} y_t^*, & y_t^* > \tau \\ \tau, & y_t^* \leq \tau \end{cases} \tag{1.1}$$

$$y_t^* = \beta x_t + u_t \tag{1.2}$$

where $\beta \in \mathbb{R}^{1\times n}$ is vector of constants, $x_t \in \mathbb{R}^{n\times 1}$ is the input vector at time t, y_t is a scalar output, and u_t is a Gaussian random variable with zero mean and variance, σ_u^2. The value u_t captures the measurement noise and is independent of both y_t and x_t. The use of ordinary least squares to estimate β or σ_u from the output would be inconsistent because the entire population of the dependent variable is not being observed.

Many methods have been devised to solve for the parameters of the Tobit model, including Tobin's original maximum likelihood estimator [2]. An analysis of this method and the consistency of the estimates is presented in [14]. In the engineering community, the Tobit model for measurements is referred to as a Wiener model. Being a linear model followed by a static nonlinearity, the Wiener model is often used to model sensor nonlinearities [15]. In [16], a recursive identification for a linear model with known saturation nonlinearity is presented. However, when the output is censored, the identification update is halted because the input has driven the output to a "slope zero" region. If the recursive estimation is to converge quickly, the input must be chosen to avoid this region.

The goal of the Tobit model is to estimate parameters of statistical data without bias introduced by conventional methods. There exist many more methods to identify Tobit model parameters and compute expectations of censored data sequences. [1, 14, 17–21]. These methods require knowledge of the entire measurement history. A recursive estimator, such as the Kalman filter, has not previously been developed for this type of measurement nonlinearity.

As discussed, censoring occurs often in engineering, science and social science; and has been used to identify models given censored or truncated data. However, despite many obvious examples of censored data in estimation and tracking, such as in sensor measurements [22] or computer vision applications, the Tobit model has not received much attention in the field of signal processing or control theory. In signal processing and control theory, the Kalman filter [23] has become ubiquitous in tracking and estimation. A Kalman filter is a recursive algorithm that operates on streams of noisy input data to produce estimates of an underlying state. It is an optimal estimate under certain assumptions of the noise distribution and system dynamics. Many estimation applications where the Kalman filter is used, especially those using low cost commercial of-the-shelf sensors (COTS), the input measurement are subject to the same censoring that are found in economics via the Tobit model, and biology via the survival function. In real-time control, censoring in embedded applications frequently takes the form of limit-of-detection, occlusion region, or sensor saturation. All of these forms, are referred to as Tobit model type 1 censoring [2]. When censored measurements are introduced into a linear model fit or estimation procedure such as the standard Kalman filter, the resulting estimates of the underlying states are biased.

Much of the work in recursive estimation has been in loosening the assumptions of the Kalman filter so it can be implemented in a wider range of applications. The unscented Kalman filter (UKF), the extended Kalman filter (EKF) and the particle filter are examples and can be applied to nonlinear systems [24]. Censoring induced by a Tobit Model is a unique type of nonlinearity in that the nonlinearity is piecewise linear but the slope of the output becomes zero in the censored region. Recursive estimators have not been implemented for this unique case. Similar measurement nonlinearities have been addressed by Kalman filtering, one being the case where the measurements are intermittent. A formulation of the Kalman filter designed for the intermittent measurement case is presented in [25, 26]. This formulation reduces to a linear predictor when measurements are missing. The estimator in [25] provides the minimum state error variance filter given all past observations and arrival sequences, and is an improvement on Jump Least Square (JLS) theory [27] which gives a minimum state error variance filter assuming only the observations and the knowledge of the previous arrival. Both of these previous formulations relied on the assumption that missed measurements were uncorrelated with the state value. The problem with this solution in a censored measurement model is that the the assumptions are violated. More specifically, the measurement is correlated to the state value as is it relates to the threshold between censored and non censored regions, the measurement model and the noise. Tobit model censoring may be formulated as an intermittent measurement problem, but because the dropped measurements are correlated with the state values the result estimates of the state are biased.

One attempt to solve the censored measurement problem was presented in [28]. This formulation treats the censored and non censored measurements differently, and has a formulation that is not recursive (it requires knowledge of the entire history of censored measurements); recursion is a major motivation for using a Kalman filter. Another type of censoring studied in [29, 30] looks at the censoring in a distributed detection system. The censored data is either sent or not sent to a fusion center based on its informativeness. This work [29,30] studies a different type of censoring from what is considered here, but shows an improvement in performance by using the information present in missing values, where the probability of missing values are correlated to state values. It is proposed that by using the underlying model and the information contained in a censored measurement, the state estimate can be improved.

Another difficulty in using a Kalman filter for censored measurements is that the measurement noise is not Gaussian near the censoring region. For example, if the state variable is a constant near the censored region, noise on the measurements causes some of the measurements to be censored, and the standard Kalman filter produces a biased estimate of the state. Past work designed estimators using the likelihood function that accounts for state dependent Gaussian observation noise. The work in [31] generated an iterative Kalman filter to solve the nonlinear least square problem according to the likelihood function.

Censoring can be generalized as an output nonlinearity, and general output nonlinearities can be addressed using the EKF, the UKF or the particle filter. With censored measurements, the state-measurement equation has a sharp discontinuity at the threshold value of the censoring region, which is a problem for the EKF because the gradient does not exist at this discontinuity. The particle filter formulated for partially observed Gaussian state space models is presented in [32]. Particle filters are much more computationally expensive than an extended Kalman filter or linear Kalman filter because they require the use of a weighted set of samples called particles to generate the posteriori distribution, $p(x_k|y_{1:k})$. Furthermore, the sharp discontinuity in the measurement model for censored data means that a large number of particles are necessary to adequately model the system in this region. The UKF is a less computationally expensive approach that proves to be non-robust when the measurements are on or near censoring. The method described in this dissertation avoids the use of numerical approximation methods such as the particle filter by directly computing the relevant posteriori distributions from the censored data measurement model. The resulting filter has a similar computational burden to the standard Kalman filter, which means it can be used in computation-limited environments such as embedded systems. In this dissertation, the first formulation of the Kalman filter capable of estimating state variables from censored data without bias is presented and referred to as the Tobit Kalman filter.

1.1 Censored Data

The early work on Tobit models produced a classification scheme for censored models. These classes define how the system is censored or truncated and gives the ability to account for censoring that is dependent on other variables. The difference between censoring and truncation is that a censored measurement model provides a measurement when the measurement is in a censored region, while truncation provides no measurement when the latent variable is within a censored region. In this section, some common censoring models are developed. The Tobit model types are grouped depending on the similarities in the likelihood functions [6]. In [21], the process for constructing likelihood functions for model parameter identification with several censoring types is presented.

1.1.1 Tobit Type 1

Tobit Type 1 is referred to as the censored regression model, in the general case the model is,

$$y_t = \begin{cases} y_t^*, & \tau_l < y_t^* < \tau_h \\ \gamma_l, & y_t^* \le \tau_l \\ \gamma_h, & y_t^* \ge \tau_h \end{cases} \tag{1.3}$$

where $\tau_l \to -\infty$, is defined as the 'right censoring' case and $\tau_h \to \infty$, the 'left censoring' case.

1.1.2 Tobit Type 2

The Tobit type 2 model, also referred to as a Heckman model [33], is similar to type 1. The difference arrives in that the censoring of the latent variable is dependent on another latent variable.

$$y(2)_t = \begin{cases} y(2)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.4}$$

1.1.3 Tobit Type 3

The combination of Tobit type 1 and 2 is a Tobit type 3 model.

$$y(1)_t = \begin{cases} y(1)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.5}$$

$$y(2)_t = \begin{cases} y(2)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.6}$$

1.1.4 Tobit Type 4

$$y(1)_t = \begin{cases} y(1)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.7}$$

$$y(2)_t = \begin{cases} y(2)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.8}$$

$$y(3)_t = \begin{cases} y(3)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.9}$$

1.1.5 Tobit Type 5

$$y(2)_t = \begin{cases} y(2)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.10}$$

$$y(3)_t = \begin{cases} y(3)_t^*, & y(1)_t^* > \tau \\ \tau, & y(1)_t^* \leq \tau \end{cases} \tag{1.11}$$

See [6] for a more in depth discussion on Tobit model types, solutions and applications.

1.2 Bayesian Estimation

In the previous section, some common censored models were introduced. Here, we introduce Bayesian statistics and its relation to estimation theory and censored measurement models, see [34] for a more indepth discussion on Bayesian analysis.

Bayesian probability is a method of making inferences on distributions based on evidence from correlated values. The Bayesian branch of probability is derived from Bayes' rule, Equation 1.12, which relates conditional probabilities, to marginal probabilities. An application to the evidential probabilistic interpretation, or Bayesian interpretation, is Bayes' inference. Bayes' inference updates the probability of a hypothesis given evidence using Bayes' rule.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1.12}$$

Bayesian probability is used often in estimation theory, which uses a model and measurements to predict, smooth, interpolate information. One might design an estimator to smooth a noisy measurement, or estimate noise distributions on a signal or estimate the phase or frequency of a sinusoidal signal. For example, the estimator for a parameter θ from a measurement y might be written as,

$$\hat{\theta} = \hat{\theta}(y) \tag{1.13}$$

A loss function, $L(\theta, \hat{\theta})$ defines the deviation from the true value. The expected value of the loss function is defined as the Bayes risk. The values of $\hat{\theta}$ that minimizes the Bayes risk, $E(L(\theta, \hat{\theta}))$, is the Bayesian estimator with the $E(L(\theta, \hat{\theta}))$ taken over the prior probability distribution. The prior probability of θ is the distribution of the parameter θ before any evidence is taken into account, while the posterior probability is the distribution conditioned on some evidence . In the cases in which the posterior probability is given, $P(\theta|y)$, the estimate variance of θ might be minimized. The estimate of θ using the posterior distribution is advantageous in application, assuming $P(\theta|y)$ has a narrower distribution than $P(\theta)$.

State estimation is commonly used in the control of dynamical systems and the Kalman filter is typically the algorithm that is employed. In the next section, the Kalman filter will be derived begining using a Bayesian framework. The Bayesian derivation of the Kalman filter can be found in several sources, including [35]. The Kalman filter estimates the states of a hidden Markov model to minimize the mean squared error estimate in x_k given measurements y_k. The Kalman filter is optimal under certain assumptions, that will declared in the following sections.

1.2.1 Markov Model

Bayesian filtering and estimation [36, 37] will briefly be reviewed, along with its relation to optimal estimation. The Bayesian filter under a Markov assumption computes the state estimate $\hat{x}_k$ of the true state x_k at time k, given measurements y_k. The Markov assumption states,

$$P(x_k|y_k...y_0, x_{k-1}...x_0) = P(x_k|y_k, x_{k-1}) \tag{1.14}$$

and the conditional probability of the measurements is,

$$P(y_k|y_{k-1}...y_0, x_k...x_0) = P(y_k|x_k). \tag{1.15}$$

A graphical representation of the hidden Markov assumption (HMA) is in Figure 1.1. The HMA is crucial to recursive state estimation because the state can be estimated using only previous measurements and states.

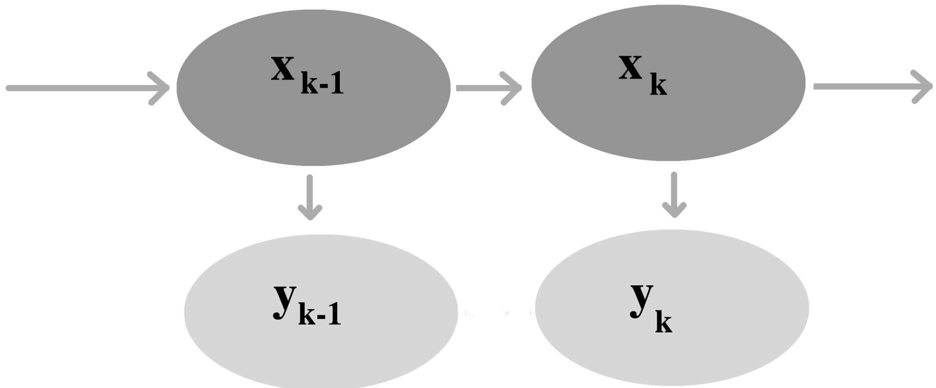

Figure 1.1: Block diagram of a hidden Markov chain

1.2.2 Sequential Bayesian Estimation

Recursive Bayesian estimation allows the Bayesian inference problem to be implemented recursively for online estimation of parameters or states given measurements. A recursive approach is necessary for control applications since the controller needs real time information of the changing parameters or states. The process for sequential Bayes estimation is to estimate the state given the process model (the state update), then to update the state estimate given measurements (the measurement update).

Using the Chapman-Kolmogorov equation [38], which provides the prior probability density function of the state, the distribution of interest is

$$P(x_k|y_{1:k-1}) = \int P(x_k|x_{k-1})P(x_{k-1}|y_{1:k-1})dx_{k-1} \tag{1.16}$$

where $P(x_k|y_{1:k-1})$ is the prior distribution and is defined by the dynamic model. This is the state update step of the filter; the measurement update is,

$$P(x_k|y_{1:k}) = \frac{P(y_k|x_k)P(x_k|y_{1:k-1})}{P(y_k|y_{1:k-1})} \tag{1.17}$$

The *likelihood* is the measurement noise model $P(y_k|x_k)$. The *evidence* is the denominator of Equation 1.17, which is

$$P(y_k|y_{1:k-1}) = \int P(y_k|x_k)P(x_k|y_{1:k-1})dx_k \tag{1.18}$$

and is often called the normalizing constant.

The *prior*, *likelihood* and the *evidence* are used to find the a *posteriori* probability. The recursive solution of Equations 1.16 and 1.17 is a conceptual solution made less complicated with the Markov assumption. However, the relationship between x_k and y_k will decide if an analytic solution is possible. If no analytic solution is possible, a sequential Monte Carlo method would work and approach the actual posterior probability density as the number of Monte Carlo samples increases. In general, these methods impose no restrictions on the model or noise distribution.

1.2.2.1 Loss Functions and Optimality Criteria

Optimal estimators are filters that meet certain criteria, for example, the minimum of a loss function. The type of loss function or any other type of criteria defines the optimality. There are several types of loss functions for estimation, in this section we will review a couple common ones. For more information on optimal estimation see [39].

In signal processing, the mean squared error (MSE) is a common measure that defines the quality of the estimate. The minimum mean squared error (MMSE) refers to an estimate that minimizes the quadratic function associated with the MSE. The cost function for the MSE is,

$$E[\|x_k - \hat{x}_k\|^2|y_k] = \int \|x_k - \hat{x}_k\|^2 P(x_k|y_k)dx_k \tag{1.19}$$

The optimality criteria in this section has the states as unknown parameters; these equations can also be formulated for parameter identification.

The properties of the MMSE are that given the measurements y_k the MMSE estimate is,

$$\hat{x}_k^{MMSE} = E(x_k|y_k) \tag{1.20}$$

The estimate is also unbiased,

$$E(\hat{x}_k^{MMSE}) = E(E(x_k|y_k)) = E(x) \tag{1.21}$$

The MSE cost function is minimized when the orthogonality principle is met, meaning the estimate error is not correlated with the measurement. Being a necessary and sufficient condition for optimality, the orthogonality principle is useful for finding the MSE estimator. In mathematical terms the orthogonality principle states,

$$\begin{aligned} E((x_k - \hat{x}_k)y_k) = 0 \\ E((x_k - \hat{x}_k)) = 0 \end{aligned} \tag{1.22}$$

In the case where x_k and y_k values are jointly Gaussian the MMSE estimate is a linear relation between the measurements and the states and can be written as,

$$E(\hat{x}_k^{MMSE}) = ay_k + b \tag{1.23}$$

This is a consequence of summation of Gaussian functions and that x_k and y_k can be collectively written as a multivariate normal distribution. Using the information that the estimator in linear and Equations 1.22, the orthogonality principle may be used to find the MMSE estimate.

Other possible optimality criteria include: Maximum a posteriori (MAP) or the mode of the posterior distribution.

$$\begin{aligned} \hat{x}_k^{MAP} &= \underset{x_k}{argmax} P(x_k|y_k) \\ &= \underset{x_k}{argmax} \frac{P(y_k|x_k)P(x_k)}{P(y_k)} \\ &= \underset{x_k}{argmax} P(y_k|x_k)P(x_k) \end{aligned} \tag{1.24}$$

where the denominator is dropped because it is independent of x_k. The advantage to using the MAP estimate is that the estimation include prior and likelihood functions, allowing the use of knowledge contained in various distributions. However, the posterior distribution may be difficult to obtain in some applications, forcing the use of Monte Carlo techniques to obtain the estimate. Even using a Monte Carlo approach to obtain the MAP estimate can cause inadequate results when the posterior distribution is multimodal. In the case where the posterior contains multiple peaks the MAP result may not produce an accurate estimate.

Another optimality criteria is the maximum likelihood estimator (MLE),

$$\hat{x}_k^{MLE} = \underset{x_k}{argmax} P(y_k|x_k) \tag{1.25}$$

When there are multiple measurements $y_{0:k}$, then

$$\hat{x}_k^{MLE} = \underset{x_k}{argmax} \prod_{k=1}^{N} P(y_k|x_k) \tag{1.26}$$

Taking the logarithm of the likelihood function, $\prod_{k=1}^{N} P(y_k|x_k)$, often provides an analytic solution for the estimate which is identical to the MLE since the logarithm is monotonically increasing.

$$\hat{x}_k^{MLE} = \underset{x_k}{argmax} \sum_{k=1}^{N} log(P(y_k|x_k)) \tag{1.27}$$

Taking the derivative of Equation 1.27 with respect to x_k, and setting the result equal to zero will find the values x_k that maximize this function.

1.2.2.2 Optimal Measurement Update

Two types of Bayesian filtering are used in practice: batch processing which uses all available data and sequential or recursive processing. The advantages to the recursive approach is that only new data is analyzed sequentially, so there is no need to store all past data. The criteria used here for recursive Bayesian filter is the MMSE, which will require infinite memory and infinite computational power when the Markov assumption cannot be made. However, if the Markov assumption is applicable a solution can be derived to obtain recursive, optimal solution. The procedure for estimating a state using a sequential Bayesian filter is to first perform a time update of the state using a prediction or smoothing model, then a measurement update using a new measurement. Let us define x_k as the state we would like to estimate, $\bar{x}_k$ is an estimate of the state using the only the time update, y_k is the measurement and $\bar{y}_k$ is the estimate of the measurement using $\bar{x}_k$. So the measurement model is as follows,

$$y_k = Cx_k \tag{1.28}$$

$$\bar{y}_k = C\bar{x}_k \tag{1.29}$$

Where $\mathbf{C} \in \mathbb{R}^{\mathbf{mxn}}$. To find the optimal estimate of x_k we use the minimum mean squared error criteria with a Gaussian assumption. The solution the the MSE criteria is the expected value of the state using the posterior probability,

$$\hat{x}_k = E[x_k|y_k] \tag{1.30}$$

And from Section 1.2.2.1 the solution to the MSE is linear when x_k and y_k are jointly Gaussian, so we know that the solution is of the form,

$$E[x_k|y_k] = \alpha y_k + \beta \tag{1.31}$$

The distributions of x_k and y_k are,

$$P(x_k) = \frac{1}{(2\pi)^{1/2}\Sigma_{xx}^{1/2}} exp(-\frac{(x_k - \bar{x}_k)}{2\Sigma_{xx}}) \tag{1.32}$$

$$P(y_k) = \frac{1}{(2\pi)^{1/2}\Sigma_{yy}^{1/2}} exp(-\frac{(y_k - \bar{y}_k)}{2\Sigma_{yy}}) \tag{1.33}$$

$$P(x_k, y_k) = \frac{1}{(2\pi)^{1/2}\Sigma_{xy}^{1/2}} exp(-\frac{(y_k - \bar{y}_k)}{2\Sigma_{xy}}) \tag{1.34}$$

Where Σ_{xx}, Σ_{yy}, and Σ_{xy}, are the covariance matrices. Using Equation 1.31 and finding the values of α and β that minimize the MSE cost function,

$$MSE(x_k) = E(x_k - E(x_k|y_k)) \tag{1.35}$$

Will yield,

$$E(x_k|y_k) = \bar{x}_k + \Sigma_{xy}\Sigma yy^{-1}(y_k - \bar{y}_k) \tag{1.36}$$

And the covariance of the measurement update,

$$cov(x_k - E(x_k|y_k)) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T \tag{1.37}$$

Where $\bar{y}_k$ and $\bar{x}_k$ are the mean values of the measurements and states respectively. This solution is the optimal estimate of x_k given that x_k and y_k are jointly Gaussian. In the non-Gaussian case Equation 1.36 provides the best linear estimate. The best linear estimate is proven with the orthogonality principle, which states that there is no other linear estimator of x_k that outperforms Equation 1.36.

1.2.3 Kalman Filter

In the previous sections we have introduced Bayesian probability theory, the recursive Bayes filter and estimation optimality. The Kalman filter deviates from the generic Bayesian interpretation by making some assumptions on the model [40]. The recursion in equations 1.16 and 1.17 will result in the Kalman filter when the noise on the measurement model and the process model are jointly Gaussian. To compute the value of $E(x_k|y_k)$ the Kalman filter computes the minimum mean squared error estimate. Thus, the Kalman filter is optimal in the mean squared sense for Gaussian distributions and is the best linear estimator when the noise is non-Gaussian. The Kalman filter assumes the following linear model, with a Markov assumption on the process equation:

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{1.38}$$

where the first equation represents the process model while the second equation represents the measurement model. The $x_k \in \mathbb{R}^{n\times 1}$ is the state vector, u_k is the known input and $y_k \in \mathbb{R}^{n\times 1}$ is the measurement vector. The $A \in \mathbb{R}^{n\times n}$ is the state transition matrix, $B \in \mathbb{R}^{1\times n}$ is the input matrix and $C \in \mathbb{R}^{1\times n}$ is the measurement state transition matrix. The w_k and v_k are Gaussian random vectors with zero mean, they have covariance $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{n\times n}$, respectively.

$$Q = E[w_k w_k^T] \tag{1.39}$$

$$R = E[v_k v_k^T] \tag{1.40}$$

the processes and measurement noise are mutually independent;

$$E[w_l v_m^T] = 0 \tag{1.41}$$

and the state and measurement noise are mutually independent;

$$E[x_l v_m^T] = 0, E[x_l w_m^T] = 0 \tag{1.42}$$

The measurement and process noise are also time independent, so,

$$E[v_l v_m^T] = \sigma_v^2 \delta(l-m), E[w_l w_m^T] = \sigma_w^2 \delta(l-m) \tag{1.43}$$

Where δ is the Kronecker delta function and σ_v^2 and σ_w^2 are the measurement and process noise covariance. In general, the correlations of states and noise can be represented as,

$$cov\left(\begin{bmatrix} w_l \\ v_l \\ x_0 \end{bmatrix}, \begin{bmatrix} w_m \\ v_m \\ x_0 \end{bmatrix}\right) = \begin{bmatrix} Q\delta(l-m) & 0 & 0 \\ 0 & R\delta(l-m) & 0 \\ 0 & 0 & P_0 \end{bmatrix} \tag{1.44}$$

Where P_0 is the initial error covariance based off the initial condition and prior. The notation for the state after a time update, the *a priori* estimate, and measurement update, the *a posteriori* estimate must be introduced. The state estimate at time k after the time update is denoted as $x_{k|k-1}$ with associated state error covariance,

$$P_{k|k-1} = E[(x_k - x_{k|k-1})[(x_k - x_{k|k-1})^T|y_{k-1}] \tag{1.45}$$

While the state estimate and state error covariance matrix is denoted as $x_{k|k}$ and $P_{k|k} = E[(x_k - x_{k|k-1})[(x_k - x_{k|k-1})^T|y_k]$. The corresponding distributions of state estimates are,

$$(x_k|y_{k-1}) \sim \mathcal{N}(x_{k|k-1}, P_{k|k-1}) \tag{1.46}$$

$$(x_k|y_k) \sim \mathcal{N}(x_{k|k}, P_{k|k}) \tag{1.47}$$

The time update step of the discrete Kalman filter is calculated using the process model in Equation 1.38,

$$x_{k|k-1} = E(x_k|y_{k-1}) = AE(x_{k-1}|y_{k-1}) + Bu_k + w_k = Ax_{k-1|k-1} + Bu_k \tag{1.48}$$

And the priori state error covariance,

$$P_{k|k-1} = cov(x_k|y_{k-1}) = AP_{k-1|k-1}A^T + Q_k \tag{1.49}$$

Where u_k has no variance because it is a known input.

Since the noise on the state and measurement are jointly Gaussian distributions, Equation 1.36 and 1.37 updates the state estimate and the state error covariance. The a posteriori estimate of the state and covariance are,

$$E(x_k|y_{k-1}) = x_{k|k-1} \tag{1.50}$$

$$cov(x_k|y_k) = P_{k|k} \tag{1.51}$$

The measurement mean and covariance,

$$E(y_k|y_{k-1}) = CE(x_k|y_{k-1}) = Cx_{k|k-1} \tag{1.52}$$

$$cov(y_k|y_{k-1}) = E(Cx_k + v_k - Cx_{k|k-1})(Cx_k + v_k - Cx_{k|k-1}) = CP_{k|k-1}C^T + R_k \tag{1.53}$$

And the cross covariance of x_k and y_k,

$$E((x_k - x_{k|k-1})(y_k - Cx_{k|k-1})^T|y_k) = CP_{k|k-1} \tag{1.54}$$

The complete measurement update is,

$$\begin{aligned} x_{k|k} &= x_{k|k-1} + CP_{k|k-1}(CP_{k|k-1}C^T + R_k)^T(y_k - E(y_k|y_k)) \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1}C^T(CP_{k|k-1}C^T + R_k)^T CP^T_{k|k-1} \end{aligned} \tag{1.55}$$

Where $E(y_k|y_k) = Cx_{k|k-1}$ and $K_k = CP_{k|k-1}(CP_{k|k-1}C^T + R_k)^T$ is the Kalman gain.

1.3 Estimation of Censored Data

The Kalman filter presented in this Chapter is not able to provide optimal or even unbiased estimates of the states when the measurements are censored. This is because the assumptions for the Kalman filter are not met when the measurement noise is censored. First, the measurement noise is not orthogonal to the state, meaning the measurement noise is correlated to the state, more so near the censored region. Secondly, the likelihood function of the censored distribution is flat when the state is censored. See Chapter 3 for the censored measurement likelihood function.

Chapter 2

CENSORED DATA IN STANDARD ESTIMATION

As discussed in Chaper 1, there exists many methods for estimation of nonlinear systems. In this chapter, the performance of these algorithms when the measurement model is the Tobit type 1 model is reviewed, but first, the performance of an estimator using a linear approach is considered.

Censoring is usually not expected or not modeled as a nonlinearity in the system. It can happen when an unusual disturbance causing sensor saturation or if a sensor is calibrated to fit the dynamic range of the data acquisition device but the measurement noise causes clipping. In these cases, a linear estimator produces biases. For example, for inertial navigation devices the inertial measurement unit (IMU) is calibrated in a laboratory environment for the conditions that are expected when used in the field [41] [42] [43]. If, for some reason the conditions change or an unexpected motion is observed with velocity, magnetic field or acceleration; the sensor may go into saturation. Motivated by this problem, we will briefly illustrate the advantage to including censoring in the dynamic model by using a Kalman filter to estimate a constant value within the dynamic range of the sensor. Consider a stationary variable with value 0.5. The measurement is subject to zero-mean Gaussian additive noise with variance of 1, left censored at 0. This is Equation 1.38 with $Cx_k = .5$ and $v_k \sim \mathcal{N}(0, 1)$.

The standard Kalman filter converges to the expected value of the censored observation y_k, or approximately 0.65, or a bias of almost 30% despite the fact that the mean value is above the censoring limit. If we consider the alternative formulation of the Kalman filter [25] [26], and treat the censored measurements as missing measurements, then the filter converges to the expected value of 1.01, for a bias of over 100%. The results of the two methods are shown in Figure 2.1.

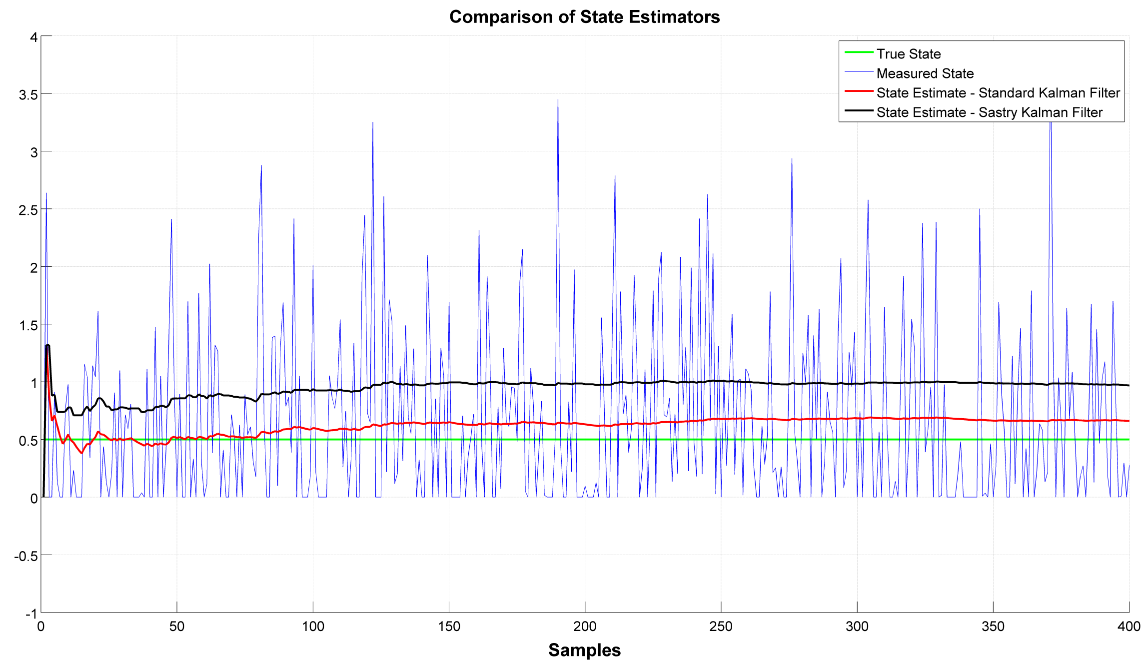

Figure 2.1: Comparison of linear state estimation with censored measurements using the Kalman filter and the Kalman filter with intermittent measurements

The bias in both of these methods would propagate into the state estimate, causing subsequent errors in a controller, or even instability. To obtain unbiased estimate of the state, the Tobit model nonlinearity must be accounted for in the estimator. There are many types of nonlinear estimator, all useful in wide ranges of applications and designs.

There are several nonlinear estimators possible with censored measurement models. Of these approaches, the most computationally expensive but accurate would be the particle filter [32]. However, the particle filter is difficult to implement when computational power is limited, as in embedded systems. Other less expensive techniques explored in this section are the EKF and the UKF [40] [44] [45] . Comparisons between the UKF and the EKF have been made in [46]. The EKF is the most widely used alternative to the Kalman filter when the model is nonlinear [47]. The EKF uses the Jacobian of the nonlinearity to update the state and measurement covariances. Convergence is not guaranteed, the EKF will diverge when the linearized transform is not sufficiently accurate, so the performance varies with model type and state location. The motivation behind creating the UKF [48–50] was to provide an alternative to the EKF which would not suffer from the EKF convergence and linearization issues. The UKF works by performing a statistical linearization around the state estimate, which is meant to improve convergence.

2.1 The Extended Kalman Filter

The discrete-time EKF starts with the model,

$$\begin{aligned} x_k &= f_{k-1}(x_k, u_{k-1}, w_{k-1}) \\ y_k &= h_k(x_k, v_k) \\ w_k &\sim \mathcal{N}(0, Q_k) \\ v_k &\sim \mathcal{N}(0, R_k) \end{aligned} \tag{2.1}$$

The f_{k-1} is assumed to be linear in our example, so $f_{k-1}(x_k, u_{k-1}, w_{k-1}) = Ax_{k-1} + w_k$, the nonlinearity is restricted to the measurement model h_k. Because of this, the state update equations of Equation 2.1 are linear, the EKF is the same as the linear Kalman filter state update equations and are optimal. See Chapter 1 for the equations of Kalman filter update.

The correction step using available measurements is,

$$\begin{aligned} K_k &= P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + M_kR_kM_k^T)^{-1} \\ x_{k|k} &= x_{k|k-1} + K_k(y_k - h_k(x_k, 0)) \\ P_{k|k} &= (I - K_kH_k)P_{k|k-1} \end{aligned} \tag{2.2}$$

With,

$$H_k = \frac{\partial h_k}{\partial x}|_{x_{k|k-1}}$$

$$M_k = \frac{\partial h_k}{\partial v}|_{x_{k|k-1}}$$

In the censored region $h(y_k^* > \tau, 0)$,

$$\frac{\partial h_k}{\partial x}\Big|_{x_{k|k-1}} = Cx_{k|k-1}$$

$$\frac{\partial h_k}{\partial v}\Big|_{x_{k|k-1}} = 1$$

In the region $h(y_k^* < \tau, 0)$,

$$\frac{\partial h_k}{\partial x}\Big|_{x_{k|k-1}} = 0$$

$$\frac{\partial h_k}{\partial v}\Big|_{x_{k|k-1}} = 0$$

and is undefined at at the threshold, τ.

2.2 The Unscented Kalman Filter

The UKF is based on the unscented transform, which is a statistical linearization approach using deterministic sigma points. The purpose of the unscented transform is to better approximate mean and covariance through nonlinear transformations. Further, in [51] [52], the argument is made that the unscented transform can approximate discontinuities in the nonlinear function, which linearized methods cannot achieve. The basic idea behind the unscented transform is that the distribution of a nonlinear function is better approximated using deterministic set of sigma points, s, than to approximate a nonlinear function by Taylor expansion. To define the UKF, the sigma points must be defined, there are $p+1$ sigma points, with weights $W(i)$ that satisfy,

$$\sum_{i=0}^{p} W(i) = 1; W(i) \geq 0 \; i = 0, 1, ..., p \tag{2.3}$$

The set of points that satisfy the above condition include,

$$\begin{aligned} &x(i) = \hat{x} + (\sqrt{N\Psi_x}) \\ &W(i) = \tfrac{1}{2N} \\ &x(i) = \hat{x} - (\sqrt{N\Psi_x}) \\ &W(i+N) = \tfrac{1}{2N} \end{aligned} \tag{2.4}$$

Where N is the length of the vector, x and Ψ_x is the covariance of x

Each point is sent through the nonlinear transform, here, h represents our measurement model in Equation 2.1.

$$y(i) = h(x(i)) \tag{2.5}$$

And the mean is taken as a weighted average,

$$\hat{y} = \sum_{i=0}^{p} W(i)y(i) \tag{2.6}$$

The covariance is calculated in a similar matter, a weighted product of the mean points corrected by the weighted average,

$$\Psi_y^{UKF} = \sum_{i=0}^{p} W(i)(y(i) - \hat{y})(y(i) - \hat{y})^T \tag{2.7}$$

In the censored measurements model, or any other type of piecewise measurement model the sigma points may cause biased results in the output mean and covariance near the discontinuity. For example, for the model,

$$\begin{aligned} y_k^* &= x_k + v_k \qquad\qquad v_k \sim \mathcal{N}(0, q) \\ y_k &= \begin{cases} y_k^*, & \tau < y_k^* \\ T, & otherwise \end{cases} \end{aligned} \tag{2.8}$$

We look at the unscented transform from x_k to y_k, where x_k is the deterministic input and y_k is the output with additive noise and censoring at a lower limit τ. We will show that the unscented transform only gives accurate measurement uncertainties when the sigma points represent the discontinuity well or when they are far from a censored region. For example, if we select sigma points such that $N = 1$, $p = 3$, the elements of $\hat{x}$ which will converge to the linear Kalman filter but be biased if $\tau < x(i_{low})$, where $x(i_{low})$ is the outermost sigma point and closest to the censored region. The effect of the bias on the measurement error covariance and expected measurement will be,

$$Bias^{UKF} = \Phi(\tau, \hat{y}, q) > \epsilon \tag{2.9}$$

With the normal distribution represented by,

$$\begin{aligned}\phi(x,\mu,\sigma) &= \tfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ \Phi(x,\mu,\sigma) &= \tfrac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x'-\mu)^2}{2\sigma^2}} dx'\end{aligned} \tag{2.10}$$

where ϵ is a small value which represents tolerable measurement bias when censoring is present [53]. In most applications where there is censoring the $Bias^{UKF}$ is always present because the normal distribution extends from $-\infty$ to $+\infty$, but the effects are minimal unless the expected measurement is near a censored region. The UKF will provide a better measurement error covariance estimate when the sigma point span the discontinuity. In Figure 2.2 we have the measurement uncertainty vs threshold limit as given by the UKF in a case where measurement noise on the latent variable $v_k \sim \mathcal{N}(0,1)$, and $x_k = 0$, according to Equation 2.8. Because the tails of the distribution are not represented, there exists large discontinuities at the locations of the sigma points around the mean of the measurement.

The bias is caused by the tails of the noise distribution, however if the discontinuity caused by censoring lies between sigma points there will also be a bias in the measurement noise covariance. This issue will be mitigated if the number of sigma point is increased, however, the estimate will not be consistent.

2.3 Particle Filter

Particle filter, or Bootstrap filtering is a sequential Monte Carlo technique that estimates the posterior density function using a Bayesian recursive framework. A set of weighted particles represent the sample distribution of the posterior density at each time step. Using the weighted particles representing the posterior distribution, an estimate can be made. The weights are the product of a recursive relation which we will derive here, it turns out that they are updated based on the likelihood of the observations given the particle location. The advantages to particle filtering is that there are no restrictions on the system or observation model or their corresponding distributions. The number of particles greatly effects the performance of the particle

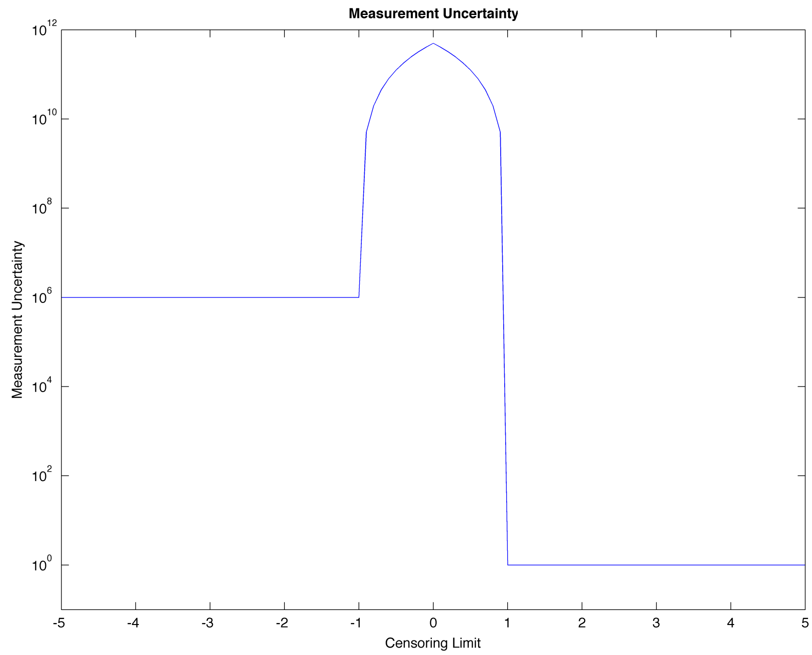

Figure 2.2: Measurement uncertainty as calculated by the UKF vs threshold limit (τ)

filter but can also be the cause of severe computation burden. As the number of particles increases the closer the estimator gets to providing the optimal estimate.

A disadvantage to the particle filter is that when applied to high dimensional systems [54], the function updating the weights tends to collapse if the dimensionality is high. This collapse of weights is an inherent problem with particle filtering and occurs when the effective sample size representing the posterior gets too small. The fix for the collapse of the sample weights is to re-sample the posteriori distribution, applying one of the many different types of resampling techniques [55]. Another issue associated with high dimensional systems is the design of the weight update function. The weighting function for a single state is influenced by all observations, even when the observations and states are mostly independent from each other. In this case, the posteriori distribution calculated by the particle filter under estimates the uncertainty in the states. To obtain good performance using a particle filter in higher order systems, the number of particles should increase exponentially with the systems size [54].

To find the recursive relation to estimate the state from a hidden Markov Model, extending from Section 1.2.1,

$$
\begin{aligned}
P(x_{0:k}|y_{1:k}) &= \frac{P(y_k|x_{0:k}, y_{1:k})P(x_{0:k}|y_{1:k})}{P(y_k|y_{1:k-1})} \\
&= \frac{P(y_k|x_{0:k}, y_{1:k-1}P(x_k|x_{0:k-1}, y_{1:k-1})P(x_{0:k-1}|y_{1:k-1})}{P(y_k|y_{1:k-1})} \\
&= \frac{P(y_k|x_k)P(x_k|x_{k-1})P(x_{0:k-1}|y_{1:k-1})}{P(y_k|y_{1:k-1})} \\
&\propto P(y_k|x_k)P(x_k|x_{k-1})P(x_{0:k-1}|y_{1:k-1})
\end{aligned}
\tag{2.11}
$$

The posterior density is represented by $\{x^i_{0:k}, w^i_{0:k}\}^{Np}_{i=1}$, which is a randomly generated set. The $x^i_{0:k}$ is the set of all states up to time k, and have weights $w^i_{0:k}$. From this random set we can approximate the posterior density to be,

$$P(x_{0:k}|y_{1:k}) \approx \sum_{i=1}^{Np} w^i_k \delta(x_{0:k} - x^i_{0:k}) \tag{2.12}$$

The weights are chosen by an importance sampling method [56]. Importance sampling is a technique for estimating statistics of an unknown distribution by using a known distribution, see [57]. The idea behind importance sampling is that regions in the probability distribution with high importance are given a higher weight. Using importance sampling the weights in Equation 2.12 are equal to,

$$w^i_k \propto \frac{P(x^i_{0:k}|y_{1:k})}{q(x^i_{0:k}|y_{1:k})} \tag{2.13}$$

Where q is the importance density and is the density we are sampling from, while the numerator in Equation 2.13 is referred to as the nominal distribution.

The distribution $P(x^i_{0:k}|y_{1:k})$, is known, let $x^i_k \approx q(x^i_{0:k}|y_{1:k}), i = 1, ..., Np$ be the samples that are generated, also known as the importance density. The importance density is factorized as follows,

$$q(x_k|y_{1:k}) = q(x_k|x_{0:k-1}, y_{1:k})q(x_{0:k-1}|y_{1:k-1}) \tag{2.14}$$

where the augmented state density function allows for $q(x_k|y_{1:k})$ to be estimated by multiplying the estimating sample distribution by the new probability distribution, $q(x_k|x_{0:k-1}, y_{1:k})$. Using Equation 2.11 and 2.14 in Equation 2.12,

$$\begin{aligned} w_k^i &\propto \frac{P(y_k|x_k^i)P(x_k^i|x_{k-1}^i)P(x_{0:k-1}|y_{1:k-1})}{q(x_k|x_{0:k-1}, y_{1:k})q(x_{0:k-1}|y_{1:k-1})} \\ &= w_{k-1}^i \frac{P(y_k|x_k)P(x_k|x_{k-1})}{q(x_k|x_{0:k-1}, y_{1:k})} \\ &= w_{k-1}^i \frac{P(y_k|x_k)P(x_k^i|x_{k-1})}{q(x_k|x_{k-1}, y_k)} \end{aligned} \tag{2.15}$$

where the weights are normalized, $\sum_i^{Np} w_k^i = 1$.

The optimal proposal distribution is,

$$q(x_k|x_{k-1}, y_k) = P(x_k|x_{k-1}, y_k) \tag{2.16}$$

where the transition prior probability obtained from the dynamical or time update model, Equation 2.17, is often used as the importance function.

$$q(x_k|x_{k-1}, y_k) = P(x_k|x_{k-1}) \tag{2.17}$$

The weight update in Equation 2.15 suffers from a degeneracy problem, which is the collapse of particle weights onto one particle. The degeneracy problem is solved by re-sampling the particles where particles with small weights are eliminated, replacing them with particles that have higher weight. This is not done every iteration, but done when the effective sample size falls below a predefined value. An estimate of the effective size is,

$$N_{eff} = \frac{1}{\sum_{i=1}^{Np}(w_k^i)^2} < N_T \tag{2.18}$$

where N_T is the predefined limit.

The three filters described in this chapter are meant to work in broad applications of nonlinear systems. The Tobit Type 1 model is a unique system in that it is linear when the measurements are not near a censoring region. As such, the censoring model can be viewed as a linear system followed by a static nonlinearity. In this application, the noise term on the measurement is subject to nonlinear transformation, so nonlinear estimator routines relying on $y_k = h(x_k) + v_k$ rather than $y_k = h(x_k, v_k)$ will not be applicable. Because the Tobit Model has an underlying linear system, it is proposed that there exists an approximate linear estimator. The next Chapter derives this filter, using a single assumption on the predictability of censoring. Alternatively, the particle filter performs the best for censoring systems though it has some disadvantages including computational and convergence restraints, that will be discussed in the simulation section. The EKF does not treat censoring as a smooth transition between regions, resulting in an undefined Jacobian at the threshold limit while the UKF is proven to be biased near the censoring region.

Chapter 3

THE TOBIT KALMAN FILTER

In this Chapter we will derive the Tobit Kalman filter by first formulating the problem as a linear system followed by the static, censoring nonlinearity. Then, the Tobit model statistics will be used to find the innovation needed for the recursive estimator, along with an analytic expression for the measurement variance. Finally, the Tobit Kalman filter gain will be derived to complete the linear estimator.

3.1 Problem Formulation

To define the censoring problem consider the evolution of a scalar output state sequence as,

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_k + w_{k-1} \\ y_k^* &= Cx_k + v_k \\ y_k &= \begin{cases} y_k^*, & y_k^* > \tau \\ \tau, & y_k^* \leq \tau \end{cases} \end{aligned} \tag{3.1}$$

$x_k \in \mathbb{R}^{n\times 1}$ is the state vector, u_k is the scalar input and y_k is the scalar measurement . The $A \in \mathbb{R}^{n\times n}$ is the state transition matrix, $B \in \mathbb{R}^{n\times 1}$ is the input matrix and $C \in \mathbb{R}^{1\times n}$ is the measurement vector. The w_k and v_k are Gaussian random vectors with zero mean, they have covariance $Q \in \mathbb{R}^{n\times n}$ and $R = \sigma^2$, respectively, where σ is the standard deviation of the measurement noise. The Kalman filter is optimal in the Gaussian sense; however when the noise distribution on y_k is a censored Gaussian, the filter is suboptimal; and since the noise is correlated to the state value, there is a violation of the assumptions of the Kalman filter. The closer the state is to the threshold value the more censored the Gaussian distribution on y_k becomes.

3.2 Problem Formulation for the Tobit Case

Using Equation 3.1, we define y_k as the censored observation and y_k^* as the latent variable. The probability distribution of a censored variable with normally distributed noise is:

$$f(y_k|x_k) = \frac{1}{\sigma}\phi(\frac{y_k - Cx_k}{\sigma})u(y_k - \tau) + \delta(\tau - y_k)\Phi(\frac{\tau - Cx_k}{\sigma}) \tag{3.2}$$

where

$$\phi(\frac{y_k - Cx_k}{\sigma}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(y_k - Cx_k)^2}{2\sigma^2}} \tag{3.3}$$

and

$$\Phi(\frac{y_k - Cx_k}{\sigma}) = \int_{-\infty}^{y_k} \frac{1}{\sqrt{2\pi}}e^{-\frac{(z_k - Cx_k)^2}{2\sigma^2}}dz_k \tag{3.4}$$

are the probability density function and the cumulative distribution function of a Gaussian random variable whose mean is Cx_k. δ is the Kronecker delta function. The $u(\alpha)$ is a step function and is equal to $u(\alpha) = 1$ when $\alpha \geq 0$ and $u(\alpha) = 0$ when $\alpha < 0$. The delta function at $\tau - y_k$ is present when measurements at the censoring limit are recorded. The delta function is absent in a Tobit type 1 model that has missing measurements when they are censored, this is referred to as the truncated model.

The likelihood function for the standard Tobit model is,

$$L = \prod_{y_k^* \leq \tau} [1 - \Phi(\frac{Cx_k - \tau}{\sigma})] \prod_{y_k^* \geq \tau} \sigma^{-1}\phi(\frac{y_k - Cx_k}{\sigma}) \tag{3.5}$$

as formulated by Tobin in his pioneering work [2].

The expected value of the measurements when uncensored is given by:

$$\begin{aligned} E(y_k|y_k > \tau, x_k, \sigma) &= \sigma^{-1}\int_{\tau}^{+\infty} z\frac{\phi(\frac{z - Cx_k}{\sigma})}{1 - \Phi(\frac{\tau - Cx_k}{\sigma})}dz \\ &= Cx_k + \sigma\lambda((\tau - Cx_k)/\sigma) \end{aligned} \tag{3.6}$$

this differs from the true value of the latent variable by a bias of $\sigma\lambda((\tau - Cx_k)/\sigma)$ where $\lambda(\alpha) = \frac{\phi(\alpha)}{[1 - \Phi(\alpha)]}$ is the inverse Mills ratio (IMR) [14].

The expected measured value when both uncensored and censored measurements are included is:

$$
\begin{aligned}
E[y_k|x_{k|k-1},\sigma] &= P[y_k > \tau|x_{k|k-1},\sigma]E[y_k|y_k > \tau, x_{k|k-1},\sigma] \\
&\quad + P[y_k = \tau|x_{k|k-1},\sigma]E[y_k|y_k = \tau, x_{k|k-1},\sigma] \\
&= \Phi(\tfrac{Cx_k-\tau}{\sigma})[Cx_k + \sigma\lambda((\tau - Cx_k)/\sigma)] + \Phi(\tfrac{\tau-Cx_k}{\sigma})\tau
\end{aligned} \tag{3.7}
$$

The variance of the expected measured value is derived in [58] and can be written as:

$$
Var[y_k|y_k > \tau, x_k, \sigma] = E[y_k^2|y_k > \tau, x_k, \sigma] - [E[y_k|y_k > \tau, x_k, \sigma]]^2 \tag{3.8}
$$

$$
E[y_k^2|y_k > \tau, x_k, \sigma] = \sigma^{-1}\frac{1}{1-\Phi(\frac{\tau-Cx_k}{\sigma})}\int_{\tau}^{+\infty} z^2\phi(\tfrac{z-Cx_k}{\sigma})dz \tag{3.9}
$$

so

$$
Var[y_k|y_k > \tau, x_k, \sigma] = \sigma^2[1 - \eth(\frac{\tau - Cx_k}{\sigma})] \tag{3.10}
$$

where

$$
\eth(\tfrac{\tau-Cx_k}{\sigma}) = \lambda(\tfrac{\tau-Cx_k}{\sigma})[\lambda(\tfrac{\tau-Cx_k}{\sigma}) - (\tfrac{\tau-Cx_k}{\sigma})] \tag{3.11}
$$

Note that $Var[y_k|x_k,\sigma] = Var[y_k|y_k > \tau, x_k, \sigma]$ since $Var[y_k|y_k < \tau, x_k, \sigma] = 0$.

3.3 Derivation of the Tobit Kalman Filter

The Bayesian derivation of the Kalman filter is found in [35]. The Bayesian filter under a Markov assumption computes the state estimate $\hat{x}_k$, of the true state x_k, at time k given measurements y_k. The Markov assumption states,

$$
P(x_k|y_k...y_0, x_{k-1}...x_0) = P(x_k|y_k, x_{k-1}) \tag{3.12}
$$

and the conditional probability of the measurements is,

$$
P(y_k|y_{k-1}...y_0, x_k...x_0) = P(y_k|x_k) \tag{3.13}
$$

For a Kalman filter we are interested in how the measurements are projected on the state estimates and future state estimates. The distribution of interest is

$$P(x_k|y_{1:k-1}) = \int P(x_k|x_{k-1})P(x_{k-1}|y_{1:k-1})dx_{k-1} \tag{3.14}$$

or is the predict step of the filter. The update state may be written as

$$P(x_k|y_{1:k}) = \frac{P(y_k|x_k)P(x_k|y_{1:k-1})}{P(y_k \mid y_{1:k-1})} \tag{3.15}$$

The recursion in equations 3.14 and 3.15 will result in the Kalman filter when the noise on the measurement model and the process model are jointly Gaussian. To compute the value of $E(x_k|y_k)$ the Kalman filter computes the minimum mean squared error estimate which is,

$$\hat{x}_k = E(x_k|y_k) = \bar{x}_k + K(y_k - \bar{y}_k) \tag{3.16}$$

where $\bar{y}_k$ and $\bar{x}_k$ are the expected values of the measurements and states respectively and K is the gain that updates state estimates with measurement error.

In the censored measurement model however the noise is a censored Gaussian in the measurement equations, resulting in the distribution given by Equation 3.2. This noise function results in a nonlinear relationship between the measurements and state values, Equation 3.16 does not hold. In the next section, we develop an alternative update schedule which recursively calculates $E(x_k|y_k)$ for Tobit censored measurements.

3.3.1 The Tobit Kalman Filter

In Section 3.3 we have reviewed the basis of the Kalman filter using Bayes' rule. In this section we derive the Kalman formulation for Tobit censored measurements. The derivation is similar to the derivation for the standard Kalman filter; however,the censoring results in new definitions for the measurement residual, and consequently for the optimal Kalman gain and the estimated state covariance. Below is the notation for our model with the state $\mathbf{x_k} \in \mathbb{R}^{n\times 1}$, and $\mathbf{y_k} \in \mathbb{R}^{m\times 1}$ being the measurement on the system.

$$\begin{aligned}\mathbf{x_k} &= \mathbf{A}\mathbf{x_{k-1}} + \mathbf{w_{k-1}} \\ \mathbf{y_k^*} &= \mathbf{C}\mathbf{x_k} + \mathbf{v_k} \\ \mathbf{y_k} &= \begin{cases} \mathbf{y_k^*}, & \mathbf{y_k^*} > \mathcal{T} \\ \mathcal{T}, & \mathbf{y_k^*} \leq \mathcal{T} \end{cases}\end{aligned} \tag{3.17}$$

The matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the state transition matrix, $\mathbf{C} \in \mathbb{R}^{m\times n}$ is the measurement model and $\mathcal{T} \in \mathbb{R}^{m\times 1}$ is vector of threshold values. The noise $\mathbf{w_k}$ and $\mathbf{v_k}$ are zero mean white Gaussian noise with covariance matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ and $\mathbf{R} \in \mathbb{R}^{m\times m}$ respectively.

3.3.2 The Predict Stage

The prior estimate of the state and it's probability distribution may be written

$$\mathbf{P}(\mathbf{x_{k|k-1}}|\mathbf{y_{k-1}}) \sim \mathcal{N}(\mathbf{E}(\mathbf{x_{k|k-1}}), \mathbf{Var}(\mathbf{x_{k|k-1}})) \tag{3.18}$$

where $\mathbf{x_{k|k-1}} \in \mathbb{R}^{n\times 1}$ is the state estimate vector of $\mathbf{x_k}$ given all estimates and measurements up to time $k-1$. The predict equation of the state may be written as

$$\mathbf{E}(\mathbf{x_{k|k-1}}|\mathbf{y_{k-1}}) = \mathbf{E}(\mathbf{A}\mathbf{x_{k-1|k-1}} + \mathbf{w_k}) = \mathbf{A}\mathbf{x_{k-1|k-1}} \tag{3.19}$$

$\mathbf{x_{k-1|k-1}}$is the estimate of $\mathbf{x_{k-1}}$. The state error covariance given measurements and state information up to time $k-1$ may be written as

$$\begin{aligned}\mathbf{cov}(\mathbf{x_{k|k-1}} - \mathbf{x_k}) &= \mathbf{cov}(\mathbf{A}\mathbf{x_{k-1|k-1}} + \mathbf{w_k} - \mathbf{A}\mathbf{x_{k-1}}) \\ &= \mathbf{A}\mathbf{Var}(\mathbf{x_{k-1|k-1}})\mathbf{A^T} + \mathbf{Q} \\ &= \mathbf{A}\boldsymbol{\Psi}_{\mathbf{k-1|k-1}}\mathbf{A^T} + \mathbf{Q}\end{aligned} \tag{3.20}$$

where $\mathbf{Q}$ is the model covariance matrix and $\boldsymbol{\Psi}_{\mathbf{k-1|k-1}}$ is the previous a posteriori estimate of the state error covariance.

3.3.3 The Update Stage

The optimal Kalman filter must minimize the state error covariance, $\mathbf{\Psi_{k|k}}$. The update stage corrects the state estimate using current measurements. The update step will reduce the state error covariance, whereas the predict step will result in a widening of the state error covariance. The Kalman correction step to obtain the current estimate given all observations up to time k may be written as

$$\mathbf{x_{k|k}} = \mathbf{x_{k|k-1}} + \mathbf{K_k}(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}|\mathbf{x_k}, \mathbf{y_{k-1}})) \tag{3.21}$$

The value of $E(y_k)$ was calculated for a scalar case for a censored value in Equation 3.7; in this notation $\mathbf{E}(\mathbf{y_k}|\mathbf{x_k}, \mathbf{y_{k-1}}) \in \mathbb{R}^{m\times 1}$ is a vector, denoted as $\mathbf{E}(\mathbf{y_k})$ for the rest of this section. Each scalar component of $\mathbf{E}(\mathbf{y_k})$ can be censored at any given time and will have different threshold limits $\mathcal{T} = [\mathcal{T}(1), \mathcal{T}(2), ..., \mathcal{T}(m)]$ with $\mathcal{T}(l)$, $y_k(l)$ representing the lth component of arrays $\mathcal{T}$ and $\mathbf{y_k}$ respectively.

$\mathbf{K_k}$ in Equation 3.21 we minimize the state error covariance,

$$\begin{aligned} \mathbf{\Psi_{k|k}} &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k}}) \\ &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k}(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}))) \end{aligned} \tag{3.22}$$

A Bernoulli random variable will be introduced to model the occurrence of a censored measurements vs an actual measurement. The variable $p_k(l) = 1$ when the measurement is not censored and $p_k(l) = 0$ when the measurement is equal to the threshold value. The measurement model can be written as

$$p_k(l) = \begin{cases} 1, & Cx_k(l) + v_k(l) > \mathcal{T}(l) \\ 0, & Cx_k(l) + v_k(l) \leq \mathcal{T}(l) \end{cases} \tag{3.23}$$

At any given time step the measurement will represent the state by $Cx_k(l)+v_k(l)$ with probability $E(p_k(l))$. In matrix notation the Bernoulli random matrix will be diagonal $\mathbf{p_k} \in \mathbb{R}^{m\times m}$ so the measurements are given by,

$$\mathbf{y_k} = \mathbf{p_k}(\mathbf{C}\mathbf{x_k} + \mathbf{v_k}) + (\mathbf{I_{m\times m}} - \mathbf{p_k})\mathcal{T} \tag{3.24}$$

where $\mathbf{I_{m\times m}}$ is is the identity matrix. Substituting into Equation 3.22 yields

$$\begin{aligned}\mathbf{\Psi_{k|k}} &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k}}) \\ &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k}(\mathbf{p_k}(\mathbf{C}\mathbf{x_k} + \mathbf{v_k}) + \\ &(\mathbf{I_{m\times m}} - \mathbf{p_k})\mathcal{T} - \mathbf{E}(\mathbf{y_k})))\end{aligned} \tag{3.25}$$

To simplify the notation in the derivation we set the Kalman error to

$$\mathbf{\tilde{y}_k} = \mathbf{p_k}(\mathbf{C}\mathbf{x_k} + \mathbf{v_k}) + (\mathbf{I_{m\times m}} - \mathbf{p_k})\mathcal{T} - \mathbf{E}(\mathbf{y_k}) \tag{3.26}$$

so the covariance of the state estimate becomes

$$\begin{aligned}\mathbf{\Psi_{k|k}} &= \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k}\mathbf{\tilde{y}_k})(\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k}\mathbf{\tilde{y}_k})^{\mathbf{T}}) \\ &= \mathbf{\Psi_{k|k-1}} - \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})\mathbf{\tilde{y}_k^T})\mathbf{K_k^T} \\ &- \mathbf{K_k}\mathbf{E}(\mathbf{\tilde{y}_k}(\mathbf{x_k} - \mathbf{x_{k|k-1}})^{\mathbf{T}}) + \mathbf{K_k}\mathbf{E}(\mathbf{\tilde{y}_k}\mathbf{\tilde{y}_k^T})\mathbf{K_k^T}\end{aligned} \tag{3.27}$$

with

$$\mathbf{\Psi_{k|k-1}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})(\mathbf{x_k} - \mathbf{x_{k|k-1}})^{\mathbf{T}}) \tag{3.28}$$

$$\mathbf{R_{\tilde{x}\tilde{y}_k}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})\mathbf{\tilde{y}_k^T}) \tag{3.29}$$

$$\mathbf{R_{\tilde{y}\tilde{y}_k}} = \mathbf{E}(\mathbf{\tilde{y}_k}\mathbf{\tilde{y}_k^T}) \tag{3.30}$$

Next, we take the trace and the derivative of equation 3.27 and set the result equal to zero to find the optimal Kalman gain.

$$\begin{aligned}\mathbf{Tr}(\mathbf{\Psi_{k|k}}) &= \mathbf{Tr}(\mathbf{\Psi_{k|k-1}}) - \mathbf{2Tr}(\mathbf{R_{\tilde{x}\tilde{y}_k}}\mathbf{K_k^T}) \\ &+ \mathbf{Tr}(\mathbf{K_k}\mathbf{R_{\tilde{y}\tilde{y}_k}}\mathbf{K_k^T}) \\ \frac{\mathbf{d}}{\mathbf{dK_k}}\mathbf{Tr}(\mathbf{\Psi_{k|k}}) &= -\mathbf{2Tr}(\mathbf{R_{\tilde{x}\tilde{y}_k}}) + \mathbf{2Tr}(\mathbf{K_k}\mathbf{R_{\tilde{y}\tilde{y}_k}}) \\ \mathbf{K_k} &= \mathbf{R_{\tilde{x}\tilde{y}_k}}\mathbf{R_{\tilde{y}\tilde{y}_k}^{-1}}\end{aligned} \tag{3.31}$$

which results in the familiar projection equation. In a standard linear Kalman filter the values of $\mathbf{R_{\tilde{x}\tilde{y}}}$ and $\mathbf{R_{\tilde{y}\tilde{y}}}$ are functions of $\mathbf{\Psi}$, $\mathbf{H}$ and $\mathbf{R}$. Because our measurements are not linearly related to the state vector in or around a censored region we must explicitly find values for $\mathbf{R_{\tilde{x}\tilde{y}}}$ and $\mathbf{R_{\tilde{y}\tilde{y}}}$. The function for $\mathbf{R_{\tilde{x}\tilde{y}}}$ is,

$$\begin{aligned}\mathbf{R_{\tilde{x}\tilde{y}_k}} &= \mathbf{E((x_k - x_{k|k-1})((Cx_k + v_k)^T p_k} + \mathcal{T}^{\mathbf{T}}\mathbf{(I_{m\times m} - p_k) - E(y_k)^T))} \\ &= \mathbf{E(x_k x_k^T C^T p_k) + E(x_k v_k^T p_k) + E(x_k} \mathcal{T}^{\mathbf{T}}\mathbf{(I_{m\times m} - p_k))} \\ &\quad \mathbf{-E(x_k)E(y_k)^T - E(x_{k|k-1} x_k^T C^T p_k) - E(x_{k|k-1} v_k^T p_k) +} \\ &\quad \mathbf{E(x_{k|k-1}} \mathcal{T}^{\mathbf{T}}\mathbf{(I_{m\times m} - p_k)) - E(x_{k|k-1})E(y_k)^T}\end{aligned} \tag{3.32}$$

The probability of the measurement being non censored is a function of the distance between the latent measured variable and the threshold value. The expected value of $p_k(l,l)$ may be written as

$$E(p_k(l,l)) = \Phi(\frac{Cx_k(l) - \mathcal{T}(l)}{\sigma(l)}) \tag{3.33}$$

Where $Cx_k(l)$ is the l^{th} element of the measurement vector and $\sigma(l)$ is the variance of the noise on that element. In principle, this requires knowledge of the true state value. The following assumption allows us to relax this dependence and use the estimated state value instead.

Assumption 1

We assume that the state prediction permits a sufficiently accurate estimate of the probability of censoring:

$$E(p_k(l,l)) = \Phi(\tfrac{Cx_k(l)-\mathcal{T}(l)}{\sigma(l)}) \approx \Phi(\tfrac{Cx_{k|k-1}(l)-\mathcal{T}(l)}{\sigma(l)}) \tag{3.34}$$

Assumption 2

In most applications the **R** matrix is diagonal, meaning the measurement noise independent amongst measurements. Because of the commonality of **R** being diagonal we will restrict our derivation to this case. Extending the derivation to a model with off-diagonal $\mathbf{R}$ elements is straight forward, but notationally cumbersome.

$$cov(y_k(d), y_k(l)) = 0 \;\forall\; d \neq l \tag{3.35}$$

3.3.4 The Update Stage, continued

The above assumptions allows us to estimate $\mathbf{p_k}$ at each iteration and obtain values of $\mathbf{R_{\tilde{x}\tilde{y}}}$ and $\mathbf{R_{ee}}$ without the knowledge of $\mathbf{x_k}$. Where Assumptions 1 and 2 hold,

$$\mathbf{E(p_k)} = \mathbf{Diag}\begin{pmatrix} \Phi(\frac{Cx_{k|k-1}(1)-\mathcal{T}(1)}{\sigma(1)}) \\ \Phi(\frac{Cx_{k|k-1}(2)-\mathcal{T}(2)}{\sigma(2)}) \\ \vdots \\ \Phi(\frac{Cx_{k|k-1}(m)-\mathcal{T}(m)}{\sigma(m)}) \end{pmatrix}. \tag{3.36}$$

Revisiting $\mathbf{R_{\tilde{x}\tilde{y}_k}}$, and using $\mathbf{E(x_{k|k-1}v_k^T)} = \mathbf{0_{n\times n}}$ since $\mathbf{v_k}$ is uncorrelated white Gaussian noise and $\mathbf{E(x_{k|k-1})} = \mathbf{x_{k|k-1}}$, $\mathbf{E(x_k)} = \mathbf{x_{k|k-1}}$ and

$$\begin{aligned} \mathbf{E(x_k x_k^T)} &= \mathbf{E((x_k - E(x_{k|k-1}))(x_k - E(x_{k|k-1}))^T)} \\ &\quad + \mathbf{E(x_k)E(x_k)^T} \\ &= \mathbf{\Psi_{k|k-1} + x_{k|k-1}x_{k|k-1}^T} \end{aligned} \tag{3.37}$$

The value of $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ is,

$$\begin{aligned} &\mathbf{R_{\tilde{x}\tilde{y}_k}} = (\mathbf{\Psi_{k|k-1} + x_{k|k-1}x_{k|k-1}^T})\mathbf{C^T E(p_k)} \\ &+\mathbf{x_{k|k-1}}\mathcal{T}^\mathbf{T}(\mathbf{I_{m\times m} - E(p_k)}) - \mathbf{x_{k|k-1}E(y_k)^T} \\ &-\mathbf{x_{k|k-1}x_{k|k-1}^T C^T E(p_k)} \\ &+\mathbf{x_{k|k-1}}\mathcal{T}^\mathbf{T}(\mathbf{I_{m\times m} - E(p_k)}) + \mathbf{x_{k|k-1}E(y_k)^T}) \\ &= \mathbf{\Psi_{k|k-1}C^T E(p_k)} \end{aligned} \tag{3.38}$$

Repeat the above steps, using the assumptions along with the definition of $E(y_k)$ to compute $\mathbf{R_{\tilde{y}\tilde{y}}}$.

$$\mathbf{R_{\tilde{y}\tilde{y}_k}} = \mathbf{E(p_k)C\Psi_{k|k-1}C^T E(p_k)} + \mathbf{E(p_k v_k v_k^T p_k)} \tag{3.39}$$

Where $\mathbf{E}(\mathbf{p_k v_k v_k^T p_k})^\mathbf{T}$ is the analytic value calculated in Equation 3.10. If Assumption 2 holds, this is a diagonal matrix written as:

$$\mathbf{E}(\mathbf{p_k v_k v_k^T p_k})^\mathbf{T} = \mathbf{Diag}\begin{pmatrix} Var[y_k(1)|x_{k|k-1}(1),\sigma(1)] \\ Var[y_k(2)|x_{k|k-1}(2),\sigma(2)] \\ \vdots \\ Var[y_k(m)|x_{k|k-1}(m),\sigma(m)] \end{pmatrix} \tag{3.40}$$

where $Var[y_k(i)|x_{k|k-1}(i),\sigma(i)]$ is calculated according to Equation 3.10. Substituting this optimal Kalman gain Equation 3.31 into Equation 3.27 yields the simplified covariance update equations:

$$\mathbf{\Psi_{k|k}} = (\mathbf{I_{m\times m}} - \mathbf{E}(\mathbf{p_k})\mathbf{K_k C})\mathbf{\Psi_{k|k-1}} \tag{3.41}$$

The complete Tobit Kalman filter is:

$$\begin{aligned} \mathbf{x_{k|k-1}} &= \mathbf{A x_{k-1|k-1}} \\ \mathbf{\Psi_{k|k-1}} &= \mathbf{A \Psi_{k-1|k-1} A^T} + \mathbf{Q} \\ \mathbf{x_{k|k}} &= \mathbf{x_{k|k-1}} + \mathbf{R_{\tilde{x}\tilde{y}_k} R^{-1}_{\tilde{y}\tilde{y}_k}}(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k})) \\ \mathbf{\Psi_{k|k}} &= (\mathbf{I_{m\times m}} - \mathbf{E}(\mathbf{p_k})\mathbf{R_{\tilde{x}\tilde{y}_k} R^{-1}_{\tilde{y}\tilde{y}_k} C})\mathbf{\Psi_{k|k-1}} \end{aligned} \tag{3.42}$$

where $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ is given by Equation 3.38, $\mathbf{R_{\tilde{y}\tilde{y}_k}}$ is given by Equation 3.39, elements of $\mathbf{E}(\mathbf{y_k})$ where we set $\mathbf{x_k} = \mathbf{x_{k|k-1}}$ is given by Equation 3.7, and $\mathbf{E}(\mathbf{p_k})$ is given by Equation 3.36.

3.3.5 Statistics of Right Censoring

Following from the derivations above, but switching from a low threshold to a high threshold will result in the following probability distribution of the measurements,

$$f(y_k|x_k) = \frac{1}{\sigma}\phi(\frac{y_k - Cx_k}{\sigma})u(\tau - y_k) + \delta(\tau - y_k)(1 - \Phi(\frac{\tau - Cx_k}{\sigma})) \tag{3.43}$$

See Figure 3.1 for the probability distribution function of a right censored measurement.

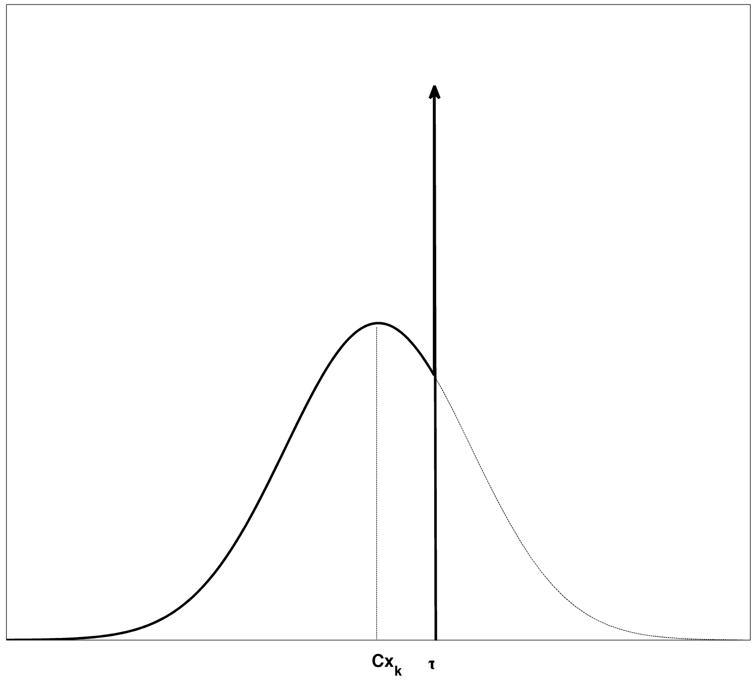

Figure 3.1: Probability distribution of a right censored measurement

The expected value of the measurements is;

$$\begin{aligned} E(y_k|y_k < \tau, x_k, \sigma) &= \sigma^{-1} \int_{-\infty}^{\tau} z \frac{\phi(\frac{z-Cx_k}{\sigma})}{\Phi(\frac{\tau-Cx_k}{\sigma})} dz \\ &= Cx_k - \sigma\lambda((Cx_k - \tau)/\sigma) \end{aligned} \tag{3.44}$$

The expected value,

$$\begin{aligned} E[y_k|x_{k|k-1}, \sigma] &= P[y_k > \tau|x_{k|k-1}, \sigma]E[y_k|y_k > \tau, x_{k|k-1}, \sigma] \\ &\quad + P[y_k = \tau|x_{k|k-1}, \sigma]E[y_k|y_k = \tau, x_{k|k-1}, \sigma] \\ &= \Phi(\tfrac{\tau - Cx_k}{\sigma})[Cx_k + \sigma\lambda((Cx_k - \tau)/\sigma)] + \Phi(\tfrac{Cx_k - \tau}{\sigma})\tau \end{aligned} \tag{3.45}$$

And the variance,

$$Var[y_k|y_k > \tau, x_k, \sigma] = \sigma^2[1 - \eth(\frac{Cx_k - \tau}{\sigma})] \tag{3.46}$$

Using the statistics presented in this section, the Tobit Kalman filter for right censoring can be implemented.

3.3.6 Statistics of Saturation Censoring

Saturation censoring will be the case when there is left and right censoring present in the measurement model. To define the left and right censoring problem we consider the evolution of a scalar output state sequence as,

$$\begin{aligned} x_k &= Ax_{k-1} + w_{k-1} \\ y_k^* &= Cx_k + v_k \\ y_k &= \begin{cases} y_k^*, & \tau_{low} < y_k^* < \tau_{high} \\ \tau_{low}, & y_k^* \leq \tau_{low} \\ \tau_{high}, & y_k^* \geq \tau_{high} \end{cases} \end{aligned} \tag{3.47}$$

It is important to note that the saturation censoring we are presenting in this dissertation is assuming that the Gaussian noise on the measurements, or the disturbance on the system, will never cause the latent variable to jump from the region $y_k^* > \tau_{high}$ to $y_k^* < \tau_{low}$ without passing through the uncensored region. One major application when the occurrence of $y_k^* > \tau_{high}$ then $y_{k+n}^* < \tau_{low}$ without a measurement in the uncensored region would occur is in a computer vision tracking application where a target in an image frame exits the field of view on one side and reenters the field of view on the other side of the frame. This is another type of censoring that is not discussed in this paper.

Using Equation 3.47, we define y_k as the saturated observation and y_k^* as the latent variable. The probability distribution of a saturated variable with normally distributed noise is:

$$\begin{aligned} f(y_k|x_k) = \frac{1}{\sigma}\phi(\frac{y_k - Cx_k}{\sigma})u(y_k - \tau_{low})u(\tau_{high} - y_k) \\ +\delta(\tau_{high} - y_k)\Phi(\frac{Cx_k - \tau_{high}}{\sigma}) + \delta(\tau_{low} - y_k)\Phi(\frac{\tau_{low} - Cx_k}{\sigma}) \end{aligned} \tag{3.48}$$

See Figure 3.2 for the probability distribution function of a saturated measurement.

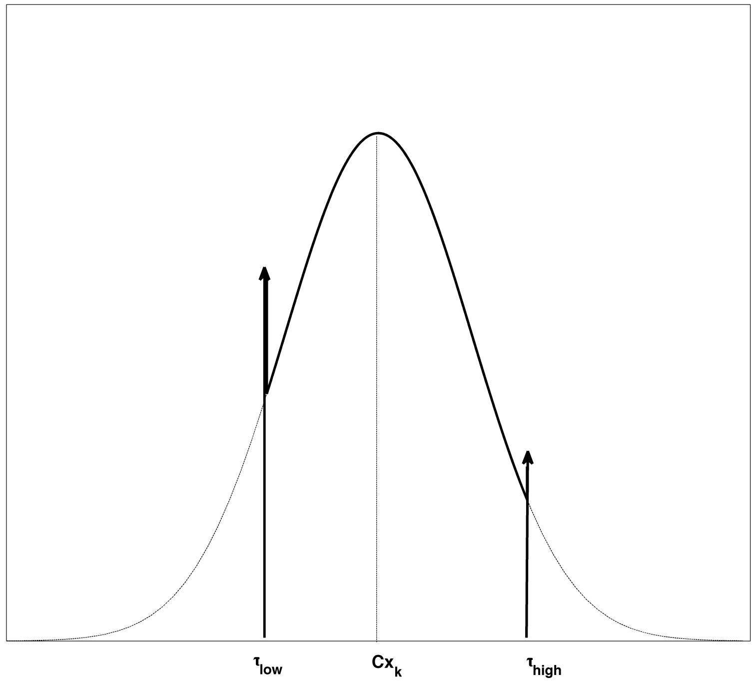

Figure 3.2: Probability distribution of a saturated measurement

To calculate the expectation of the measurements we first must find the probability of a measurement being uncensored, p_{uc}, the probability of being censored from above, p_h, and the probability of being censored from below, p_l.

$$p_{uc} = \int_{\tau_{low}}^{\tau_{high}} \phi(\frac{z - Cx_k}{\sigma})dz = \Phi(\frac{\tau_{high} - Cx_k}{\sigma}) - \Phi(\frac{\tau_{low} - Cx_k}{\sigma}) \tag{3.49}$$

$$p_{low} = \int_{-\infty}^{\tau_{low}} \phi(\frac{z - Cx_k}{\sigma})dz = \Phi(\frac{\tau_{low} - Cx_k}{\sigma}) \tag{3.50}$$

$$p_{high} = \int_{\tau_{high}}^{\infty} \phi(\frac{z - Cx_k}{\sigma})dz = 1 - \Phi(\frac{\tau_h - Cx_k}{\sigma}) \tag{3.51}$$

The mean of the measurements when the measurements are not censored is given by:

$$E(y_k|\tau_{high} > y_k > \tau_{low}, x_k, \sigma) = (\sigma p_{uc})^{-1} \int_{\tau_{low}}^{\tau_{high}} z\phi(\tfrac{z - Cx_k}{\sigma})dz \\ = Cx_k - \sigma\lambda(\tau_{high}, \tau_{low}) \tag{3.52}$$

p_{uc} is a normalization factor in the uncensored region. This expectation differs from the true value of the latent variable by a bias of $\sigma\lambda(T_h, T_l)$ where,

$$\lambda(\tau_{high}, \tau_{low}) = \frac{\phi(\frac{\tau_{high} - Cx_k}{\sigma}) - \phi(\frac{T_{low} - Cx_k}{\sigma})}{p_{uc}} \tag{3.53}$$

The expected measured value when censored measurements are included is:,

$$E[y_k|x_{k|k-1}, \sigma] = P[\tau_{high} > y_k > \tau_{low}]E[y_k|\tau_{high} > y_k > \tau_{low}] \\ +P[y_k < \tau_{low}]E[y_k|y_k < \tau_{low}] \\ +P[\tau_{high} < y_k]E[y_k|\tau_{high} < y_k] \tag{3.54}$$

In the above expectations and probability distribution functions the $x_{k|k-1}, \sigma$ is dropped for notational purposes, we write $P(\alpha|x_{k|k-1}, \sigma)$ as $P(\alpha)$. The remaining two expectations are $E[y_k|y_k < \tau_{low}] = \tau_{low}$ and $E[y_k|\tau_{high} < y_k]] = \tau_{high}$

The variance of the expected measured value is derived below:

$$Var[y_k|\tau_{high} > y_k > \tau_{low}] = \\ E[y_k^2|\tau_{high} > y_k > \tau_{low}] - [E[y_k|\tau_{high} > y_k > \tau_{low}]]^2 \tag{3.55}$$

With the first term being,

$$E[y_k^2|\tau_{high} > y_k > \tau_{low}] = \sigma^{-1}\frac{1}{p_{uc}} \int_{\tau_{low}}^{\tau_{high}} z^2\phi(\tfrac{z - Cx_k}{\sigma})dz \\ = (Cx_k)^2 + \sigma^2 - \sigma Cx_k\lambda(\tau_{high}, \tau_{low}) \\ + \frac{\sigma\tau_{low}\phi(\frac{\tau_{high} - Cx_k}{\sigma}) - \tau_{high}\phi(\frac{\tau_h - Cx_k}{\sigma})}{p_{uc}} \tag{3.56}$$

Note that $Var[y_k|x_k, \sigma] = Var[y_k|\tau_{high} > y_k > \tau_{low}]$ since $Var[y_k|y_k < \tau_{low}, x_k, \sigma] = Var[y_k|y_k > \tau_{high}, x_k, \sigma] = 0$.

3.4 The Tobit Kalman Filter for Saturation

In the previous section we have defined the statistics of a saturated measurement. In this section we derive the optimal Kalman formulation for Tobit censored measurements using a linear estimator, resulting in a predict and update stage of the Tobit Kalman filter for saturated data. The derivation of a Tobit Kalman filter for saturation is different than the left or right censoring estimator because the measurements cannot be modeled with a single Bernoulli random variable.

3.4.1 The Predict Stage

The prior estimate of the state and it's probability distribution may be written

$$\mathbf{P}(\mathbf{x_{k|k-1}}) \sim \mathcal{N}(\mathbf{E}(\mathbf{x_{k|k-1}}), \mathbf{Var}(\mathbf{x_{k|k-1}})) \tag{3.57}$$

where $\mathbf{x_{k|k-1}} \in \mathbb{R}^{n\times 1}$ is the state estimate vector of $\mathbf{x_k}$ given all estimates and measurements up to time $k-1$. The predict equation of the state may be written as

$$\mathbf{E}(\mathbf{x_{k|k-1}}) = \mathbf{E}(\mathbf{A}\mathbf{x_{k-1|k-1}} + \mathbf{w_k}) = \mathbf{A}\mathbf{x_{k-1|k-1}} \tag{3.58}$$

$\mathbf{x_{k-1|k-1}}$is the estimate of $\mathbf{x_{k-1}}$. The state error covariance given measurements and state information up to time $k-1$ may be written as

$$\begin{aligned}\mathbf{cov}(\mathbf{x_{k|k-1}} - \mathbf{x_k}) &= \mathbf{cov}(\mathbf{A}\mathbf{x_{k-1|k-1}} + \mathbf{w_k} - \mathbf{A}\mathbf{x_{k-1}}) \\ &= \mathbf{A}\mathbf{cov}(\mathbf{x_{k-1|k-1}} - \mathbf{x_{k-1}})\mathbf{A^T} + \mathbf{Q} \\ &= \mathbf{A}\boldsymbol{\Psi}_{\mathbf{k-1|k-1}}\mathbf{A^T} + \mathbf{Q}\end{aligned} \tag{3.59}$$

where $\mathbf{Q}$ is the model covariance matrix and $\boldsymbol{\Psi}_{\mathbf{k-1|k-1}}$ is the previous a posteriori estimate of the state error covariance. Again, the predict stage remains the same as the standard Kalman filter.

3.4.2 The Update Stage

The optimal Kalman filter must minimize the state error covariance, $\boldsymbol{\Psi}_{\mathbf{k|k}}$. The update step shown below will minimize the state error covariance,

$$\mathbf{x_{k|k}} = \mathbf{x_{k|k-1}} + \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}_\mathbf{k}}\mathbf{R}^{-1}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}_\mathbf{k}}(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k})) \tag{3.60}$$

This is a linear estimator that minimizes the mean squared error, as seen in Chapter 1. To obtain our estimate of $\mathbf{x_{k|k}}$ we must have the values of $\mathbf{y_k}$, $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ and $\mathbf{R_{\tilde{y}\tilde{y}_k}}$ where $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ is the cross covariance between the Kalman error and the state and $\mathbf{R_{\tilde{y}\tilde{y}_k}}$ is the variance of the Kalman error. Both $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ and $\mathbf{R_{\tilde{y}\tilde{y}_k}}$ will be defined in the upcoming paragraphs.

$$\mathbf{R_{\tilde{x}\tilde{y}_k}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}))^{\mathbf{T}}) \tag{3.61}$$

$$\mathbf{R_{\tilde{y}\tilde{y}_k}} = \mathbf{E}((\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}))(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}))^{\mathbf{T}}) \tag{3.62}$$

The value of $\mathbf{E}(\mathbf{y_k})$ was calculated for a scalar case for a censored value in Equation 3.54; in this notation $\mathbf{E}(\mathbf{y_k}) \in \mathbb{R}^{m\times 1}$ is a vector, in which each scalar component can be censored at any given time and may have different threshold limits $\mathcal{T}_{\mathbf{high}} = [\mathcal{T}_{high}(1), \mathcal{T}_{high}(2), ..., \mathcal{T}_{high}(m)]$, $\mathcal{T}_{\mathbf{low}} = [\mathcal{T}_{low}(1), \mathcal{T}_{low}(2), ..., \mathcal{T}_{low}(m)]$ with $\mathcal{T}_{high}(l)$, $\mathcal{T}_{low}(l)$, $y_k(l)$ representing the lth component of arrays $\mathcal{T}_{\mathbf{high}}$, $\mathcal{T}_{\mathbf{low}}$ and $\mathbf{y_k}$ respectively.

To find $\mathbf{K_k} = \mathbf{R_{\tilde{x}\tilde{y}_k}} \mathbf{R^{-1}_{\tilde{y}\tilde{y}_k}}$ in Equation 3.60 we minimize the state error covariance,

$$\begin{aligned} \mathbf{\Psi_{k|k}} &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k}}) \\ &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k}(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}))) \end{aligned} \tag{3.63}$$

Three Bernoulli random variables (ζ,ξ, ν) will be introduced to model the occurrence of a censored measurements at $\mathcal{T}_{high}$, $\mathcal{T}_{low}$ and a measurement of the latent variable when it is in the uncensored region respectively. The variable $\zeta_k(l) = 1$ when the measurement is censored at $\mathcal{T}_{high}$ and $\zeta_k(l) = 0$ when the measurement is not equal to the threshold value. The measurement model for the Bernoulli variables are,

$$\zeta_k(l) = \begin{cases} 1, & Cx_k(l) + v_t(l) > \mathcal{T}_{high}(l) \\ 0, & Cx_k(l) + v_t(l) \leq \mathcal{T}_{high}(l) \end{cases} \tag{3.64}$$

$$\xi_k(l) = \begin{cases} 1, & Cx_k(l) + v_t(l) < \mathcal{T}_{low}(l) \\ 0, & Cx_k(l) + v_t(l) \geq \mathcal{T}_{low}(l) \end{cases} \tag{3.65}$$

$$\nu_k(l) = \begin{cases} 1, & T_{low}(l) < Cx_k(l) + v_t(l) < \mathcal{T}_{high}(l) \\ 0, & otherwise \end{cases} \tag{3.66}$$

At any given time step the measurement will represent the state by $Cx_k(l)+v_k(l)$ with probability $E(\nu_k(l))$. In matrix notation the Bernoulli random matrices will be diagonal $\zeta_{\mathbf{k}} \in \mathbb{R}^{m\times m}$, $\xi_{\mathbf{k}} \in \mathbb{R}^{m\times m}$,$\nu_{\mathbf{k}} \in \mathbb{R}^{m\times m}$so the measurements will be arriving by the following equation

$$\mathbf{y_k} = \nu_{\mathbf{k}}(\mathbf{Cx_k} + \mathbf{v_k}) + \zeta_{\mathbf{k}}\mathcal{T}_{\mathbf{high}} + \xi_{\mathbf{k}}\mathcal{T}_{\mathbf{low}} \tag{3.67}$$

Substituting into Equation 3.63 yields

$$\begin{aligned} \mathbf{\Psi_{k|k}} &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k}}) \\ &= \mathbf{cov}(\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k}(\nu_{\mathbf{k}}(\mathbf{Cx_k} + \mathbf{v_k}) \\ &\quad +\zeta_{\mathbf{k}}\mathcal{T}_{\mathbf{high}} + +\xi_{\mathbf{k}}\mathcal{T}_{\mathbf{low}})) \end{aligned} \tag{3.68}$$

To simplify the notation in the derivation we set the Kalman error to

$$\mathbf{\tilde{y}_k} = \nu_{\mathbf{k}}(\mathbf{Cx_k} + \mathbf{v_k}) + \zeta_{\mathbf{k}}\mathcal{T}_{\mathbf{high}} + \xi_{\mathbf{k}}\mathcal{T}_{\mathbf{low}} - \mathbf{E}(\mathbf{y_k}) \tag{3.69}$$

so the covariance of the state estimate becomes

$$\begin{aligned} &\mathbf{\Psi_{k|k}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k\tilde{y}_k}) \\ &(\mathbf{x_k} - \mathbf{x_{k|k-1}} - \mathbf{K_k\tilde{y}_k})^{\mathbf{T}}) \\ &= \mathbf{\Psi_{k|k-1}} - \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})\mathbf{\tilde{y}_k^T})\mathbf{K_k^T} \\ &-\mathbf{K_kE}(\mathbf{\tilde{y}_k}(\mathbf{x_k} - \mathbf{x_{k|k-1}})^{\mathbf{T}}) + \mathbf{K_kE}(\mathbf{G_k\tilde{y}_k^T})\mathbf{K_k^T} \end{aligned} \tag{3.70}$$

with

$$\mathbf{\Psi_{k|k-1}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})(\mathbf{x_k} - \mathbf{x_{k|k-1}})^{\mathbf{T}}) \tag{3.71}$$

$$\mathbf{R_{\tilde{x}\tilde{y}_k}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})\mathbf{\tilde{y}_k^T}) \tag{3.72}$$

$$\mathbf{R_{\tilde{y}\tilde{y}_k}} = \mathbf{E}(\mathbf{\tilde{y}_k \tilde{y}_k^T}) \tag{3.73}$$

Now we need to find the values for $\mathbf{R_{\tilde{x}\tilde{y}}}$ and $\mathbf{R_{\tilde{y}\tilde{y}}}$. The function for $\mathbf{R_{\tilde{x}\tilde{y}}}$ is,

$$\begin{aligned}
&\mathbf{R_{\tilde{x}\tilde{y}_k}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})(\nu_\mathbf{k}(\mathbf{Cx_k} + \mathbf{v_k}) \\
&+\zeta_\mathbf{k}\mathcal{T}_\mathbf{high} + \xi_\mathbf{k}\mathcal{T}_\mathbf{low} - \mathbf{E}(\mathbf{y_k}))^\mathbf{T}) \\
&= \mathbf{E}(\mathbf{x_k}(\mathbf{Cx_k} + \mathbf{v_k})^\mathbf{T}\nu_\mathbf{k}^\mathbf{T}) + \mathbf{E}(\mathbf{x_k}\mathcal{T}_\mathbf{high}^\mathbf{T}\zeta_\mathbf{k}^\mathbf{T}) \\
&+\mathbf{E}(\mathbf{x_k}\mathcal{T}_\mathbf{low}^\mathbf{T}\xi_\mathbf{k}^\mathbf{T}) - \mathbf{E}(\mathbf{x_k}\mathbf{E}(\mathbf{y_k})^\mathbf{T}) \\
&-\mathbf{E}(\mathbf{x_{k|k-1}}(\mathbf{Cx_k} + \mathbf{v_k})^\mathbf{T}\nu_\mathbf{k}^\mathbf{T}) - \mathbf{E}(\mathbf{x_{k|k-1}}\mathcal{T}_\mathbf{high}^\mathbf{T}\zeta_\mathbf{k}^\mathbf{T}) \\
&-\mathbf{E}(\mathbf{x_{k|k-1}}\mathcal{T}_\mathbf{low}^\mathbf{T}\xi_\mathbf{k}^\mathbf{T}) + \mathbf{E}(\mathbf{x_{k|k-1}}\mathbf{E}(\mathbf{y_k})^\mathbf{T})
\end{aligned} \tag{3.74}$$

The probability of the measurement being non censored is a function of the distance between the latent measured variable and the threshold value. The expected value of $\zeta_k(l,l)$, $\xi_k(l,l)$ and $\nu_k(l,l)$ may be written as

$$E(\zeta_k(l,l)) = \Phi(\frac{Cx_k(l) - \mathcal{T}_{high}(l)}{\sigma(l)}) \tag{3.75}$$

$$E(\xi_k(l,l)) = \Phi(\frac{\mathcal{T}_{low}(l) - Cx_k(l)}{\sigma(l)}) \tag{3.76}$$

$$E(\nu_k(l,l)) = \Phi(\frac{\mathcal{T}_{high}(l) - Cx_k(l)}{\sigma(l)}) - \Phi(\frac{\mathcal{T}_{low}(l) - Cx_k(l)}{\sigma(l)}) \tag{3.77}$$

Where $Cx_k(l)$ is the l^{th} element of the measurement vector and $\sigma(l)$ is the variance of the noise on that element. The above equation require knowledge of the true state value, the following assumption allows us to relax this dependence and use the estimated state value instead to obtain values for $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ and $\mathbf{R_{\tilde{y}\tilde{y}_k}}$.

Assumption 1

We assume that the state prediction is a sufficiently accurate estimate of the probability of censoring from above or below:

$$E(\zeta_k(l,l)) = \Phi(\frac{Cx_k(l) - \mathcal{T}_{high}(l)}{\sigma(l)}) \approx \Phi(\frac{Cx_{k|k-1}(l) - \mathcal{T}_{high}(l)}{\sigma(l)}) \tag{3.78}$$

$$E(\xi_k(l,l)) = \Phi(\frac{\mathcal{T}_{low}(l) - Cx_k(l)}{\sigma(l)}) \approx \Phi(\frac{\mathcal{T}_{low}(l) - Cx_{k|k-1}(l)}{\sigma(l)}) \tag{3.79}$$

$$\begin{aligned} E(\nu_k(l,l)) &= \Phi(\tfrac{\mathcal{T}_{high}(l)-Cx_k(l)}{\sigma(l)}) - \Phi(\tfrac{\mathcal{T}_{low}(l)-Cx_k(l)}{\sigma(l)}) \\ &\approx \Phi(\tfrac{\mathcal{T}_{high}(l)-Cx_{k|k-1}(l)}{\sigma(l)}) - \Phi(\tfrac{\mathcal{T}_{low}(l)-Cx_{k|k-1}(l)}{\sigma(l)}) \end{aligned} \tag{3.80}$$

Assumption 2

For simplicity we assume no cross-dependence in the measurements. Consequently, $\mathbf{R}$ is diagonal and:

$$cov(y_k(d), y_k(l)) = 0 \ \forall \ d \neq l \tag{3.81}$$

3.4.3 The Update Stage, Continued

The above assumptions allows us to estimate $\zeta_\mathbf{k}$, $\xi_\mathbf{k}$ and $\nu_\mathbf{k}$ at each iteration and obtain values of $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}}$ and $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ without the knowledge of $\mathbf{x_k}$. Where Assumptions 1 and 2 hold,

$$\mathbf{E}(\zeta_\mathbf{k}) = \mathbf{Diag}\begin{pmatrix} \Phi(\frac{Cx_{k|k-1}(1)-\mathcal{T}_{high}(1)}{\sigma(1)}) \\ \Phi(\frac{Cx_{k|k-1}(2)-\mathcal{T}_{high}(2)}{\sigma(2)}) \\ \vdots \\ \Phi(\frac{Cx_{k|k-1}(m)-\mathcal{T}_{high}(m)}{\sigma(m)}) \end{pmatrix}. \tag{3.82}$$

$$\mathbf{E}(\xi_\mathbf{k}) = \mathbf{Diag}\begin{pmatrix} \Phi(\frac{\mathcal{T}_{low}(1)-Cx_{k|k-1}(1)}{\sigma(1)}) \\ \Phi(\frac{\mathcal{T}_{low}(2)-Cx_{k|k-1}(2)}{\sigma(2)}) \\ \vdots \\ \Phi(\frac{\mathcal{T}_{low}(m)-Cx_{k|k-1}(m)}{\sigma(m)}) \end{pmatrix}. \tag{3.83}$$

$$\mathbf{E}(\nu_\mathbf{k}) = \mathbf{Diag}\begin{pmatrix} \Phi(\frac{\mathcal{T}_{high}(1)-Cx_{k|k-1}(1)}{\sigma(1)}) - \Phi(\frac{\mathcal{T}_{low}(1)-Cx_{k|k-1}(1)}{\sigma(1)}) \\ \Phi(\frac{\mathcal{T}_{high}(2)-Cx_{k|k-1}(2)}{\sigma(2)}) - \Phi(\frac{\mathcal{T}_{low}(2)-Cx_{k|k-1}(2)}{\sigma(2)}) \\ \vdots \\ \Phi(\frac{\mathcal{T}_{high}(m)-Cx_{k|k-1}(m)}{\sigma(m)}) - \Phi(\frac{\mathcal{T}_{low}(m)-Cx_{k|k-1}(m)}{\sigma(m)}) \end{pmatrix}. \tag{3.84}$$

Revisiting $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}_\mathbf{k}}$, and using $\mathbf{E}(\mathbf{x}_{\mathbf{k|k-1}}) = \mathbf{x}_{\mathbf{k|k-1}}$, $\mathbf{E}(\mathbf{x}_\mathbf{k}) = \mathbf{x}_{\mathbf{k|k-1}}$ and

$$\begin{aligned} \mathbf{E}(\mathbf{x}_\mathbf{k}\mathbf{x}_\mathbf{k}^\mathbf{T}) = \mathbf{E}((\mathbf{x}_\mathbf{k} - \mathbf{E}(\mathbf{x}_{\mathbf{k|k-1}}))(\mathbf{x}_\mathbf{k} - \mathbf{E}(\mathbf{x}_{\mathbf{k|k-1}}))^\mathbf{T}) \\ +\mathbf{E}(\mathbf{x}_\mathbf{k})\mathbf{E}(\mathbf{x}_\mathbf{k})^\mathbf{T} \\ = \mathbf{\Psi}_{\mathbf{k|k-1}} + \mathbf{x}_{\mathbf{k|k-1}}\mathbf{x}_{\mathbf{k|k-1}}^\mathbf{T} \end{aligned} \tag{3.85}$$

$$\begin{aligned} &\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}_\mathbf{k}} = (\mathbf{\Psi}_{\mathbf{k|k-1}} + \mathbf{x}_{\mathbf{k|k-1}}\mathbf{x}_{\mathbf{k|k-1}}^\mathbf{T})\mathbf{C}^\mathbf{T}\mathbf{E}(\nu_\mathbf{k}) \\ &+\mathbf{x}_{\mathbf{k|k-1}}\mathcal{T}_{\mathbf{high}}^\mathbf{T}\mathbf{E}(\zeta_\mathbf{k}) + \mathbf{x}_{\mathbf{k|k-1}}\mathcal{T}_{\mathbf{low}}^\mathbf{T}\mathbf{E}(\xi_\mathbf{k}) \\ &-\mathbf{x}_{\mathbf{k|k-1}}\mathbf{x}_{\mathbf{k|k-1}}^\mathbf{T}\mathbf{C}^\mathbf{T}\mathbf{E}(\nu_\mathbf{k})^\mathbf{T} - \mathbf{x}_{\mathbf{k|k-1}}\mathcal{T}_{\mathbf{high}}^\mathbf{T}\mathbf{E}(\zeta_\mathbf{k})^\mathbf{T} \\ &-\mathbf{x}_{\mathbf{k|k-1}}\mathcal{T}_{\mathbf{low}}^\mathbf{T}\xi_\mathbf{k}^\mathbf{T} \\ &= \mathbf{\Psi}_{\mathbf{k|k-1}}\mathbf{C}^\mathbf{T}\mathbf{E}(\nu_\mathbf{k}) \end{aligned} \tag{3.86}$$

Repeat the above steps for $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}}$ to compute $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$

$$\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}_\mathbf{k}} = \mathbf{E}(\nu_\mathbf{k})\mathbf{C}\mathbf{\Psi}_{\mathbf{k|k-1}}\mathbf{C}^\mathbf{T}\mathbf{E}(\nu_\mathbf{k})+\mathbf{E}(\nu_\mathbf{k}\mathbf{v}_\mathbf{k}\mathbf{v}_\mathbf{k}^\mathbf{T}\nu_\mathbf{k}) \tag{3.87}$$

where $\mathbf{E}(\nu_\mathbf{k}\mathbf{v}_\mathbf{k}\mathbf{v}_\mathbf{k}^\mathbf{T}\nu_\mathbf{k})^\mathbf{T}$ is related to the scalar Equation 3.55. With assumption 2, the diagonal matrix is written as:

$$\mathbf{E}(\nu_\mathbf{k}\mathbf{v}_\mathbf{k}\mathbf{v}_\mathbf{k}^\mathbf{T}\nu_\mathbf{k})^\mathbf{T} =$$

$$\mathbf{Diag}\begin{pmatrix} Var[y_k(1)|\mathcal{T}_{high}(1) > y_k(1) > \mathcal{T}_{low}(1)] \\ Var[y_k(2)|\mathcal{T}_{high}(2) > y_k(2) > \mathcal{T}_{low}(2)] \\ \vdots \\ Var[y_k(m)|\mathcal{T}_{high}(m) > y_k(m) > \mathcal{T}_{low}(m)] \end{pmatrix} \tag{3.88}$$

where $Var[y_k(i)|\mathcal{T}_{high}(i) > y_k(i) > \mathcal{T}_{low}(i)]$ is calculated according to Equation 3.55. Substituting this optimal Kalman gain into Equation 3.70 yields the simplified covariance update equations:

$$\mathbf{\Psi_{k|k}} = (\mathbf{I_{m\times m}} - \mathbf{K_k E}(\nu_\mathbf{k})\mathbf{C})\mathbf{\Psi_{k|k-1}} \tag{3.89}$$

The complete Tobit Kalman filter for saturated data is:

$$\begin{aligned} \mathbf{x_{k|k-1}} &= \mathbf{A x_{k-1|k-1}} \\ \mathbf{\Psi_{k|k-1}} &= \mathbf{A \Psi_{k-1|k-1} A^T} + \mathbf{Q} \\ \mathbf{x_{k|k}} &= \mathbf{x_{k|k-1}} + \mathbf{R_{\tilde{x}\tilde{y}_k} R^{-1}_{\tilde{y}\tilde{y}_k}}(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k})) \\ \mathbf{\Psi_{k|k}} &= (\mathbf{I_{m\times m}} - \mathbf{R_{\tilde{x}\tilde{y}_k} R^{-1}_{\tilde{y}\tilde{y}_k} E}(\nu_\mathbf{k})\mathbf{C})\mathbf{\Psi_{k|k-1}} \end{aligned} \tag{3.90}$$

where $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ is given by Equation 3.86, $\mathbf{R_{\tilde{y}\tilde{y}_k}}$ is given by Equation 3.87, $\mathbf{E}(\mathbf{y_k})$ is given by Equation 3.54, and $\mathbf{E}(\nu_\mathbf{k})$ is given by Equation 3.84.

Chapter 4

PROPERTIES OF THE TOBIT KALMAN FILTER

In this Chapter we will show some properties of the Tobit Kalman filter. We'll show that the Tobit Kalman filter converges to the Kalman filter when the effect of censoring is negligible. In addition, the small added computational burden is addressed along with a comparison of the covariance estimates to other methods.

4.1 Equivalence to the Standard Kalman Filter

The Tobit Kalman filter will converge to the standard Kalman filter when the state value is far away from the censoring region, in the one sided case,

$$\lim_{\frac{\mathbf{x_{k|k-1}}-\mathcal{T}}{\sigma}\to\infty}\begin{cases}\mathbf{E}(\mathbf{p_k}) = \mathbf{I_{m\times m}}\\ \mathbf{E}(\mathbf{y_k}) = \mathbf{C}\mathbf{x_{k|k-1}}\\ \mathbf{R} = \sigma^{\mathbf{2}}\\ \mathbf{R_{\tilde{x}\tilde{y}}} = \mathbf{C}\boldsymbol{\Psi}_{\mathbf{k|k-1}}\\ \mathbf{R_{\tilde{y}\tilde{y}}} = \mathbf{C}\boldsymbol{\Psi}_{\mathbf{k|k-1}}\mathbf{C^T} + \mathbf{R}\\ \boldsymbol{\Psi}_{\mathbf{k|k}} = (\mathbf{I_{m\times m}} - \mathbf{K_k}\mathbf{C})\boldsymbol{\Psi}_{\mathbf{k|k-1}}\end{cases} \tag{4.1}$$

so the Tobit Kalman filter is a generalization of a standard Kalman filter.

4.2 Computation of the Tobit Kalman Filter

There are a few computational advantages to using the Tobit Kalman filter over existing methods, the first being that there is only a small increase in computations compared to the standard Kalman filter. The computational difference between the Kalman filter and the Tobit Kalman filter is the addition of $2\times m$ normal PDFs and $2\times m$ normal CDFs. The extra computations are performed only once per iteration of the estimator, and are only needed for the update stage.

4.3 Comparisons of Transformations, Estimating Measurement Uncertainty

In this section we will compare the estimated measurement error covariance of the UKF and the Tobit Kalman filter. Point clouds with known distribution will be generated in two dimensions to display the unscented and Tobit transforms to the measurement domain. To facilitate language used in this section the state error covariance will be denoted by Ψ and Ψ^{UKF}, the measurement error covariance $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ for the Tobit Kalman filter and the latent measurement error covariance of the Tobit Kalman filter, denoted by Σ.

4.3.1 Statistics of Spatial Region Censoring

The distribution of the latent measurements is modeled as a 2D multivariate Gaussian,

$$\phi(x_k, y_k) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x_k-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y_k-\mu_y}{\sigma_y}\right)^2} \tag{4.2}$$

The μ_x and μ_y are the mean of of the measurements in the horizontal and vertical directions.

The censored distribution is portrayed in Figure 4.1, mathematically we have,

$$\phi_c(x_k, y_k, \Sigma_x, \Sigma_y) = \begin{cases} \frac{1}{2\pi\Sigma_x\Sigma_y} e^{-\frac{1}{2}\left(\frac{x_k-\mu_x}{\Sigma_x}\right)^2 + \left(\frac{y_k-\mu_y}{\Sigma_y}\right)^2}, & x_k, y_k \notin O \\ T, & otherwise \end{cases} \tag{4.3}$$

Where O represents the occlusion we cannot measure within. The coordinates of are bounded by $O \in \{O_{x_{max}}, O_{x_{min}}, O_{y_{max}}, O_{y_{min}}\}$

To perform Tobit Kalman measurement covariance we need $\mathbf{P_{uc}}$ for the 2D model which is,

$$\mathbf{P_{uc}} = \left(1 - \int_{x_k, y_k \notin O} \phi_c(x_k, y_k, \Sigma_x, \Sigma_y) dx_k dy_k\right) \mathbf{I_{2x2}} \tag{4.4}$$

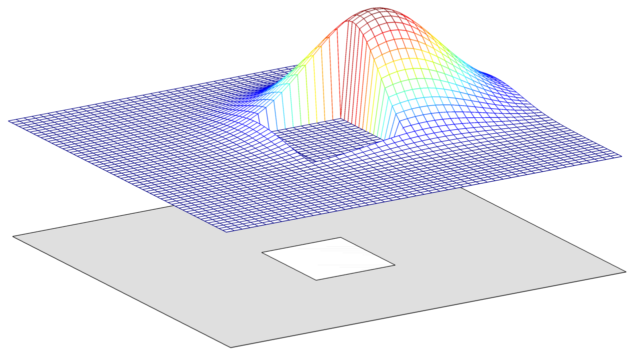

Figure 4.1: 2D probability distribution of measurement with single occlusion

Where $\mathbf{I_{2x2}}$ is the identity matrix with size $2x2$ and,

$$\begin{aligned} &\int_{x_k,y_k \notin O} \phi_c(x_k, y_k, \Sigma_x, \Sigma_y) dx_k dy_k = 1- \\ &(\Phi(O_{x_{max}}, \mu_x, \Sigma_x) - \Phi(O_{x_{min}}, \mu_x, \Sigma_x)) \\ &(\Phi(O_{y_{max}}, \mu_x, \Sigma_y) - \Phi(O_{y_{min}}, \mu_x, \Sigma_y)) \end{aligned} \tag{4.5}$$

Where Σ_x and Σ_y are the latent measurement uncertainty. To compute the latent measurement uncertainty, the same method is used that computed the measurement uncertainty. To start, the latent measurement equation for 2D is,

$$\mathbf{y_k^*} = \mathbf{Cx_k} + \mathbf{v_k} \tag{4.6}$$

With $\mathbf{C} = \mathbf{I_{2x2}}$, $\mathbf{x_k} = [x_k \ y_k]$ with x_k, y_k representing coordinates in the horizontal and vertical directions. An $\mathbf{v_k}$ is additive noise vector with covariance $\mathbf{R}$.

To calculate the uncertainty we will compute the covariance,

$$\begin{aligned} &\mathbf{E}((\mathbf{y_k^*} - \mathbf{\hat{y}_k^*})(\mathbf{y_k^*} - \mathbf{\hat{y}_k^*})^\mathbf{T}) \\ &= \mathbf{E}((\mathbf{Cx_k} + \mathbf{v_k} - \mathbf{C\hat{x}_k})(\mathbf{Cx_k} + \mathbf{v_k} - \mathbf{C\hat{x}_k})^\mathbf{T}) \\ &= \mathbf{C\Psi C^T} + \mathbf{R} \end{aligned} \tag{4.7}$$

The value for the measurement covariance as calculated by the Tobit Kalman filter is derived in the Chapter 3. To obtain the covariance in the 2D case we will use the moment generating function,

$$\mathbf{M}_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u^p v^q f(u,v) du dv$$

We will notate all elements of $\mathbf{R_k}$ below, where $\mathbf{R_k}$ is the dynamic measurement covariance for the Tobit Kalman filter. $\mathbf{R_k}$ has off diagonal elements in a spatial occlusion model because the censored region in the x direction is dependent on the censored region in the y direction.

The second moment in the x direction,

$$\mathbf{M}_{20|UC} = \mathbf{M}_{20} - \left(\mathbf{V}(x, \mu, \sigma, O_{x_{max}}, O_{x_{min}}) \left(\Phi \left(\frac{O_{y_{max}} - \mu_y}{\sigma} \right) - \Phi \left(\frac{O_{y_{min}} - \mu_y}{\sigma} \right) \right) \right) \tag{4.8}$$

Where $\mathbf{M}_{20} = \mathbf{R}(1,1)$ is the measurement variance without the occlusion. The second moment in the y direction is,

$$\mathbf{M}_{02|UC} = \mathbf{M}_{02} - \left(\mathbf{V}(y, \mu, \sigma, O_{y_{max}}, O_{y_{min}} \left(\Phi \left(\frac{O_{x_{max}} - \mu_x}{\sigma} \right) - \Phi \left(\frac{O_{x_{min}} - \mu_x}{\sigma} \right) \right) \right) \tag{4.9}$$

The nature of the 2D example is that there is a heavy dependence between the x_k and y_k variables, this will be apparent in the cross terms of the measurement noise matrix $\mathbf{R_k}$. The cross dependence between x_k and y_k is,

$$\mathbf{M}_{11|UC} = \mathbf{M}_{11} - \left(\mathbf{H}(x, \mu_x, \sigma, O_{x_{max}}, O_{x_{min}}) \mathbf{H}(y, \mu_y, \sigma, O_{y_{max}}, O_{y_{min}}) \right) \tag{4.10}$$

Where,

$$\begin{aligned} \mathbf{V}(x, \mu, \sigma, a, b) &= \frac{1}{\sigma} \int_b^a x'^2 \phi \left(\frac{x' - \mu}{\sigma} \right) dx' \\ &= \left(\left(\frac{\mu}{\sigma} + 1 \right) \Phi \left(\frac{a - \mu}{\sigma} \right) - \left(\frac{\mu + 1}{\sigma} \right) \phi \left(\frac{a - \mu}{\sigma} \right) \right) \\ &\quad - \left(\left(\frac{\mu}{\sigma} + 1 \right) \Phi \left(\frac{b - \mu}{\sigma} \right) - \left(\frac{\mu + 1}{\sigma} \right) \phi \left(\frac{b - \mu}{\sigma} \right) \right) \end{aligned} \tag{4.11}$$

$$\begin{aligned}\mathbf{H}(x,\mu,\sigma,a,b) &= \frac{1}{\sigma}\int_b^a x'\phi\left(\frac{x'-\mu}{\sigma}\right)dx' \\ &= \left(\mu\Phi\left(\frac{a-\mu}{\sigma}\right)-\sigma\phi\left(\frac{a-\mu}{\sigma}\right)\right)-\left(\mu\Phi\left(\frac{b-\mu}{\sigma}\right)-\sigma\phi\left(\frac{b-\mu}{\sigma}\right)\right)\end{aligned} \tag{4.12}$$

The final solution for the $\mathbf{R_k}$ is,

$$\mathbf{R_k} = \begin{bmatrix} \mathbf{M}_{20|UC} - \mathbf{M}^2_{10|UC} & \mathbf{M}_{11|UC} - \mathbf{M}_{10|UC}\mathbf{M}_{01|UC} \\ \mathbf{M}_{11|UC} - \mathbf{M}_{10|UC}\mathbf{M}_{01|UC} & \mathbf{M}_{02|UC} - \mathbf{M}^2_{01|UC} \end{bmatrix}$$

Where $\mathbf{M}_{10|UC}$ and $\mathbf{M}_{10|UC}$ are the expected values of the measurements in the uncensored region, and are given by,

$$\mathbf{M}_{10|UC} = \mathbf{M}_{10} - \left(\mathbf{H}(x,\mu,\sigma,O_{n\ x_{max}},O_{n\ x_{min}}\left(\Phi\left(\frac{O_{n\ y_{max}}-\mu_y^k}{\sigma}\right)-\Phi\left(\frac{O_{n\ y_{min}}-\mu_y^k}{\sigma}\right)\right)\right)$$

The first moment in the y direction is,

$$\mathbf{M}_{01|UC} = \mathbf{M}_{10} - \left(\mathbf{H}(y,\mu,\sigma,O_{n\ y_{max}},O_{n\ y_{min}}\left(\Phi\left(\frac{O_{n\ x_{max}}-\mu_x^k}{\sigma}\right)-\Phi\left(\frac{O_{n\ x_{min}}-\mu_x^k}{\sigma}\right)\right)\right)$$

Using the above equations and the Tobit Kalman filter we have a measurement uncertainty of,

$$\begin{aligned}\mathbf{R_{\tilde{y}\tilde{y}_k}} &= E((y_k-\hat{y_k})(y_k-\hat{y_k})^T) \\ &\approx \mathbf{P_{uc}C\Psi_{k|k-1}C^TP_{uc}} + +\mathbf{R_k}\end{aligned} \tag{4.13}$$

In Kalman filtering there is the state error covariance, the measurement error covariance and the cross covariance between the state and the measurement error. Similar recursive equations have been derived in [53], but with the addition of latent measurement error covariance, derived in Equation 4.7.

In Figure 4.2 we have a simulation comparing the input x_k distribution with the output distribution y_k. The input is centered at $(x_k, y_k) = (0, 0)$ and has a noise distribution of $\mathcal{N}(0,1)$ in both directions. The measurements are censored in the right half plane after a measurement noise is added of $\mathcal{N}(0,1)$ in both directions. As we can see in Figure 4.2 the Tobit measurement uncertainty, denoted by the black ellipsoid is correctly skewed to the right, consistent with the computed measurement skew from the Monte Carlo trial. By contrast, the UKF incorrectly skews to the left.

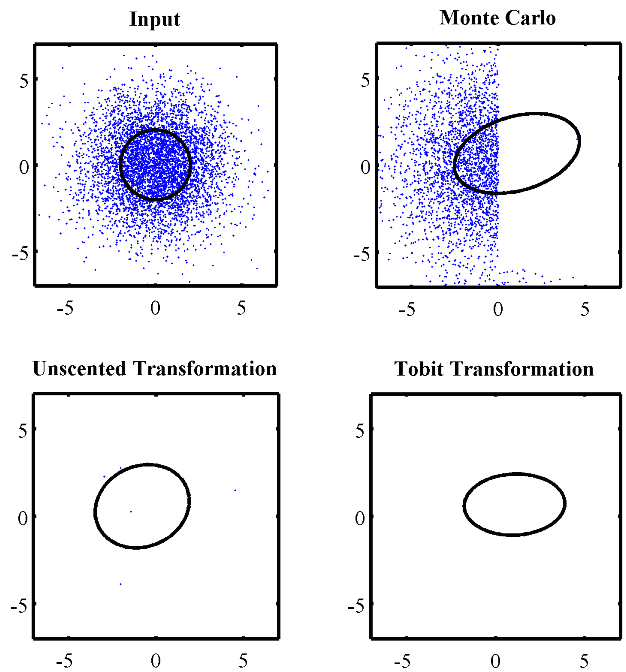

Figure 4.2: Simulation of measurement noise distribution as calculated by the Tobit Kalman update and the UKF in 2D

Chapter 5

THE 2D TOBIT KALMAN FILTER

As stated in Chapter 1, censored data is prevalent in engineering applications, one application is in computer vision. Censoring in computer vision occurs when targets or objects that need to be detected or tracked are partially or fully occluded. Censoring occurs when tracked targets exit a frame (frame censoring) and if a target enters or goes behind an occlusion (occlusion censoring). The Kalman filter for target tracking using computer vision has become possible given the availability of efficient and cheaper computing platforms with equally improved high quality cameras. The Kalman filter has been used in feature and target tracking, stereo vision [59] and consequently robot localization [60].

The applications using tracking with computer vision include tracking targets with a static background [61] [62], automated surveillance and tracking from aerial imagery [63] [64], automated video tagging for high amounts of video data [65], contour tracking for partially occluded targets [66, 67] human computer interaction through hand gestures [68], monitoring tasks such as traffic cameras [69], navigation for unmanned aerial vehicles [64], robotics [70, 71], and road vehicles. The tracking process in a vision system consists of detection of the object of interest then subsequent tracking from frame to frame using the assumption that the object will not change appearance completely in the next frame [59]. Most tracking tasks will impose this constraint on motion and appearance of the object to make the correspondence problem possible to solve. The correspondence problem is the ability to detect an object frame to frame. The algorithm for tracking should still be immune to small deformations, illuminations changes, noise in image, partial and full occlusion. Some successful work has been done in tracking that relies on the assumption that the object is deforming slowly with respect to the frame rate [61]. The Tobit Kalman for 2D tracking presented in this chapter will not focus on detection, but on tracking and subsequently imposing constraints on the motion by using an appropriate state space model. Tracking with known motion models often will allow detection algorithms to not be as complex, which is advantageous when video frames get dropped, there are missed detections or the target appearance changes [62].

There has been much work done in tracking through occlusion, in this case object tracking algorithms will first disregard pixels that are occluded [72] to perform the correspondence. Two other techniques for object tracking through occlusion in video sequences is presented in [73], where the merge split routine and the straight-through approach are presented. The correspondence problem, especially after some form of censoring in computer vision tasks remain to be difficult challenges for a tracking system.

The advantages to using a Kalman filter for computer vision based object tracking is that spurious measurements caused by inaccurate detections are suppressed due to the use of a dynamic model of the objects motion. As stated in Chapter 1, the Kalman filter performs a prediction step that provides *a priori* estimate of the state, then an update step that provides *a posteriori* estimate of the state using measurements on the system. If a measurement is not available, in a computer vision system this happens when an object of interest is not detected in a frame, then the Kalman filter will not be able to perform an update stage. Consequently, the predict stage always increases the uncertainty in the state estimate causing the state error covariance to grow if there continues to be no measurement for many sequential frames. Advantages to using Kalman filters for tracking include the ability to track and filter out spurious or noisy measurements but to also track an estimate of the state error covariance. The state error covariance should intuitively grow the longer an object is not detected in a tracking system, however; it should not grow unbounded when the measurement has entered an occlusion with known dimensions. In computer vision tasks, especially in surveillance when the motion of a target can be identified and modeled appropriately, the censored measurements when the target is behind an occlusion can still provide information to the estimator. If a target has entered a known occlusion, the state error covariance should not grow unbounded as if the measurements are missing [25, 26]. In the surveillance application, if a target is tracked as it enters a building or tunnel, we will continue to propagate the model until the target should exit on the other side. If the target fails to exit the occlusion, then the estimate should converge to the center of the occlusion.

In this chapter we wish to advance the Kalman filter to work in areas of 1D and 2D occlusions. Much like in [40], where constraints on a Kalman filter are implemented in the filter to gain performance and in some case preserve optimality. The derivation for a Tobit Kalman fitler for occlusion censoring starts off with an adapted state space model, which has piecewise format to show discontinuities of censoring. Then, a statistical approach is used to find new values for the expected measurement and variance of measurement noise to perform appropriate measurement updates for censored data. In effect, we are increasing the system knowledge by giving occlusion boundaries and standard deviation of measurements to propagate our state estimates through occlusions.

5.1 Problem Formulation in the 1D Case

In this section we will present the Tobit statistics for 1D occlusions, also referred to as dead-zones in sensors. First, we solve the problem in the case of one occlusion.

$$\mathbf{x_k} = \mathbf{A}\mathbf{x_{k-1}} + \mathbf{v_k}$$
$$\mathbf{y_k^*} = \mathbf{C}\mathbf{x_k} + \mathbf{w_k}$$
$$\mathbf{y_k} = \begin{cases} \mathbf{y_k^*}, & (\mathcal{T}_{low} > \mathbf{C}\mathbf{x_k} + \mathbf{w_k}) \cup (\mathbf{C}\mathbf{x_k} + \mathbf{w_k} > \mathcal{T}_{high}) \\ \frac{\mathcal{T}_{high} - \mathcal{T}_{low}}{2}, & otherwise \end{cases}$$

Where $\mathbf{x_k} \in \mathbb{R}^{nx1}$ is the state vector, $\mathbf{y_k} \in \mathbb{R}^{mx1}$ is the measurement vector, $\mathbf{v_k} \in \mathbb{R}^{nx1}$ with covariance matrix $\mathbf{Q}$ and $\mathbf{w_k} \in \mathbb{R}^{mx1}$ with covariance matrix $\mathbf{R}$ are the additive noise in the process and measurements respectively. The bounds for the occluded region for each measurement is $\mathcal{T}_{low} \in \mathbb{R}^{mx1}$ and $\mathcal{T}_{low} \in \mathbb{R}^{mx1}$. If the lth

measurement censoring bounds are, $\mathcal{T}_{low}(l) = \mathcal{T}_{low}(l)$, then the lth measurement does not contain an occluded region.

If there are N occlusions, then the censored measurement depends on which object occludes the target. Thus,

$$\begin{aligned}
\mathbf{x_k} &= \mathbf{A}\mathbf{x_{k-1}} + \mathbf{v_k} \\
\mathbf{y_k^*} &= \mathbf{C}\mathbf{x_k} + \mathbf{w_k} \\
\mathbf{y_k} &= \begin{cases}
\mathbf{y_k^*}, & \mathbf{C}\mathbf{x_k} + \mathbf{w_k} > \mathcal{T}_{high,1} \\
M(\mathcal{T}_{high,1}, \mathcal{T}_{low,1}), & \mathcal{T}_{low,1} < \mathbf{C}\mathbf{x_k} + \mathbf{w_k} < \mathcal{T}_{high,1} \\
\mathbf{y_k^*}, & \mathcal{T}_{high,2} < \mathbf{C}\mathbf{x_k} + \mathbf{w_k} < \mathcal{T}_{low,1} \\
M(\mathcal{T}_{high,2}, \mathcal{T}_{low,2}), & \mathcal{T}_{low,2} < \mathbf{C}\mathbf{x_k} + \mathbf{w_k} < \mathcal{T}_{high,2} \\
\vdots & \\
\mathbf{y_k^*}, & \mathbf{C}\mathbf{x_k} + \mathbf{w_k} < \mathcal{T}_{low,N}
\end{cases}
\end{aligned}$$

where $M(\mathcal{T}_{high,\eta}, \mathcal{T}_{low,\eta})$ is the midpoint between $\mathcal{T}_{high,\eta}$ and $\mathcal{T}_{low,\eta}$ and $\mathcal{T}_{high} \in \mathbb{R}^{mx1}$ and $\mathcal{T}_{low} \in \mathbb{R}^{mx1}$ are vectors of threshold limits for each element in $\mathbf{y_k}$.

5.2 Statistics of the Measurement Noise Model

The quantity $f(\mathbf{y}|\mathbf{R})$ represents the multivariate normal distribution,

$$f(\mathbf{y_k}|\mathbf{R}) = \frac{1}{\sqrt{|\mathbf{R}|(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{y_k}-\mu_\mathbf{y})\mathbf{R}^{-1}(\mathbf{y_k}-\mu_\mathbf{y})}$$

Where $\mu_\mathbf{y}$ is the mean of $\mathbf{y_k}$. With no censoring, $\mathcal{T}_{low} =\rightarrow -\infty$ and $\mathcal{T}_{high} \rightarrow \infty$, and the noise $\mathbf{w_k}$ and $\mathbf{v_k}$ are zero-mean, normally distributed, the mean values of the states and measurements models are,

$$\begin{aligned}
\mu_{\mathbf{x,uncensored}} &= \mathbf{E}(\mathbf{x_k}) = \mathbf{E}(\mathbf{A}\mathbf{x_{k-1}} + \mathbf{v_k}) = \mathbf{A}\mathbf{x_{k-1}} \\
\mu_{\mathbf{y,uncensored}} &= \mathbf{E}(\mathbf{y_k^*}) = \mathbf{E}(\mathbf{C}\mathbf{x_k} + \mathbf{w_k}) = \mathbf{C}\mathbf{x_k}
\end{aligned} \tag{5.1}$$

When censoring occurs, $(\mathcal{T}_{low,\eta} \neq \mathcal{T}_{high,\eta}) \cup (|\mathcal{T}_{high,\eta}| < \infty) \cup (|\mathcal{T}_{low,\eta}| < \infty)$, the mean value for the measurement equation given in Equation 5.1 is biased, but the expected value can be computed given *a priori* knowledge of the standard deviation of the measurement noise in the uncensored case, and an unbiased estimate of $\mathbf{x_k}$. Similarly, in the censoring case, the covariance of the measurement noise, $\mathbf{R_k}$ changes with respect to the distance the measurements are from the censoring limit. In the next sections we will refer to $\mathbf{R}$ as the covariance matrix in the non censoring case, and $\mathbf{R_k}$ the dynamic covariance matrix in censoring situations.

5.3 The Tobit Kalman Filter

The Tobit Kalman filter will be used to provide unbiased estimate of the states. The Tobit Kalman filter has many advantages when the measurents are censored or may become censored, one advantage is that when measurments are censored the error covariance matrix $\mathbf{\Psi_{k|k}} = \mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k}})(\mathbf{x_k} - \mathbf{x_{k|k}})^\mathbf{T})$ does not grow as rapidly as it does if you treat the censored measurements as missing. Below is the Tobit Kalman filter,

$$\begin{aligned}
\mathbf{x_{k|k-1}} &= \mathbf{A}\mathbf{x_{k-1|k-1}} \\
\mathbf{\Psi_{k|k-1}} &= \mathbf{A}\mathbf{\Psi_{k-1|k-1}}\mathbf{A^T} + \mathbf{Q} \\
\mathbf{x_{k|k}} &= \mathbf{x_{k|k-1}} + \mathbf{R_{\tilde{x}\tilde{y}_k}}\mathbf{R^{-1}_{\tilde{y}\tilde{y}_k}}(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k})) \\
\mathbf{\Psi_{k|k}} &= (\mathbf{I_{m\times m}} - \mathbf{R_{\tilde{x}\tilde{y}_k}}\mathbf{R^{-1}_{\tilde{y}\tilde{y}_k}}\mathbf{p_{uc}}\mathbf{C})\mathbf{\Psi_{k|k-1}}
\end{aligned} \tag{5.2}$$

$$\mathbf{R_{\tilde{x}\tilde{y}_k}} = \mathbf{\Psi_{k|k-1}}\mathbf{C^T}\mathbf{p_{uc}} \tag{5.3}$$

$$\mathbf{R_{\tilde{y}\tilde{y}_k}} = \mathbf{p_{uc}}\mathbf{C}\mathbf{\Psi_{k|k-1}}\mathbf{C^T}\mathbf{p_{uc}} + \mathbf{R_k} \tag{5.4}$$

Where $\mathbf{x_{k|k-1}}$ and $\mathbf{\Psi_{k|k-1}}$ are the *a priori* estimate of the state and state error covariance matrix, and $\mathbf{x_{k|k}}$, $\mathbf{\Psi_{k|k}}$ are the *a posteriori* estimate of the state and covariance matrix at time k. The equations $\mathbf{R_{\tilde{x}\tilde{y}_k}}$ and $\mathbf{R_{\tilde{y}\tilde{y}_k}}$ are derived in [53] and Chapter 3 and represent the covariance between the Kalman error and the state error $\mathbf{E}((\mathbf{x_k} - \mathbf{x_{k|k-1}})(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}|\mathbf{x_{k|k-1}}))^\mathbf{T})$, and the variance of the Kalman error, $\mathbf{E}((\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}))(\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}))^\mathbf{T})$. In this formulation, the Tobit Kalman filter gain is

$\mathbf{K_k} = \mathbf{R_{\tilde{x}\tilde{y}_k}} \mathbf{R^{-1}_{\tilde{y}\tilde{y}_k}}$ and the Tobit innovation is $\mathbf{y_k} - \mathbf{E}(\mathbf{y_k}|\mathbf{x_{k|k-1}})$. $\mathbf{E}(\mathbf{y}|\mathbf{x_{k|k-1}})$ is different for different censoring conditions, in the subsequent sections we will derive this term for our occlusion model. In Section 5.4 we derive $\mathbf{E}(\mathbf{y}|\mathbf{x_{k|k-1}})$ and $\mathbf{R_k}$ for one dimensional occlusions then extend the model to 2D in Section 5.7.

5.4 Statistics of the 1D Occlusion Model, Continued

Using the Tobit Kalman filter we have an estimate for $\mathbf{y_k}$, which we will denote as $\mathbf{E}(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}})$ which is unbiased mean of the measurements.

In the censored case with N occlusions the probability distribution of the measurement noise is,

$$f_c(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) = \begin{cases} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}), & \mathbf{y^*_k} > \mathcal{T}_{high,1} \\ \Delta_{k,1} \int_{\mathcal{T}_{low,1}}^{\mathcal{T}_{high,1}} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} & \mathcal{T}_{low,1} < \mathbf{y^*_k} < \mathcal{T}_{high,1} \\ f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}), & \mathcal{T}_{high,2} < \mathbf{y^*_k} < \mathcal{T}_{low,1} \\ \Delta_{k,2} \int_{\mathcal{T}_{low,2}}^{\mathcal{T}_{high,2}} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} & \mathcal{T}_{low,2} < \mathbf{y^*_k} < \mathcal{T}_{high,2} \\ \vdots & \\ f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) & \mathbf{y^*_k} < \mathcal{T}_{low,N} \end{cases} \tag{5.5}$$

With $\Delta_{k,1} = \delta(\mathbf{y_k} - M(\mathcal{T}_{high,1}, \mathcal{T}_{low,1}))$, the Kronecker delta function and $\mathcal{T}_{high,\eta}$ and $\mathcal{T}_{low,\eta}$ represent vector of bounds of the ηth occlusion. See Figure 5.1 for a representation of this distribution with $N = 3$.

Using the distribution in Equation 5.5, we can calculate the expected value of the measurements and $\mathbf{R_k}$ given the *a priori* state estimate and the uncensored measurement noise, $\mathbf{R}$.

The distribution in Equation 5.5 is valid in the general case, when there is no uncertainty in that state estimate. However, in adaptive filtering and estimation there is always some level of uncertainty. Fortunately, the Kalman filter and Tobit Kalman filter provide an estimate of the state, $\mathbf{x_{k|k}}$ and an estimate of the state error covariance matrix, $\mathbf{\Psi_{k|k}}$.

Using $\mathbf{\Psi_{k|k-1}}$ and $\mathbf{R}$, we can adjust the probability distribution to give more accurate descriptions of the probability of being censored in certain regions. The new covariance matrix that will represent the total uncertainty is $\mathbf{\Pi_k}$.

$$\begin{aligned}\mathbf{\Pi_k} &= \mathbf{E}((\mathbf{Cx_k} + \mathbf{w_k} - \mathbf{Cx_{k|k-1}})(\mathbf{Cx_k} + \mathbf{w_k} - \mathbf{Cx_{k|k-1}})^{\mathbf{T}}) \\ &= \mathbf{C\Psi_{k|k-1}C^T} + \mathbf{R}\end{aligned}$$

So the new distribution with N occlusions will be,

$$f_c(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}) =$$
$$\begin{cases} f(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}), & \mathbf{y_k^*} > \mathcal{T}_{high,1} \\ \Delta_{k,1}\int_{\mathcal{T}_{low,1}}^{\mathcal{T}_{high,1}} f(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}})d\mathbf{y_k} & \mathcal{T}_{low,1} < \mathbf{y_k^*} < \mathcal{T}_{high,1} \\ f(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}), & \mathcal{T}_{high,2} < \mathbf{y_k^*} < \mathcal{T}_{low,1} \\ \Delta_{k,2}\int_{\mathcal{T}_{low,2}}^{\mathcal{T}_{high,2}} f(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}})d\mathbf{y_k} & \mathcal{T}_{low,2} < \mathbf{y_k^*} < \mathcal{T}_{high,2} \\ \vdots & \\ f(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}) & \mathbf{y_k^*} < \mathcal{T}_{low,N} \end{cases}$$

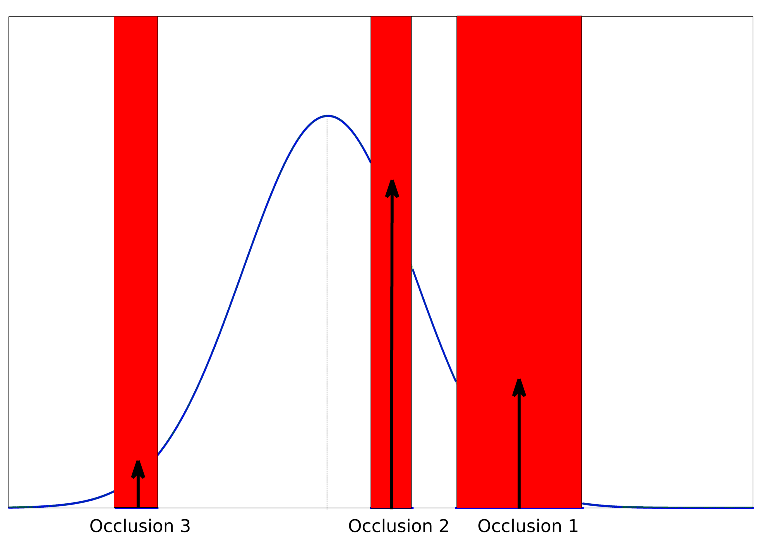

Figure 5.1: 1D distribution represented in blue with N=3 occlusions

The expected value of the measurements in all regions is,

$$
\begin{aligned}
&E(\mathbf{y_k}|\mathbf{y_k^*} > \mathcal{T}_{high,1}) = \int_{\mathcal{T}_{high,1}}^{\infty} \mathbf{y_k} f_c(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} \\
&E(\mathbf{y_k}|\mathcal{T}_{low,1} < \mathbf{y_k^*} < \mathcal{T}_{high,1}) = M(\mathcal{T}_{low,1}, \mathcal{T}_{high,1}) \\
&E(\mathbf{y_k}|\mathcal{T}_{high,2} < \mathbf{y_k^*} < \mathcal{T}_{low,1}) = \int_{\mathcal{T}_{high,2}}^{\mathcal{T}_{low,1}} \mathbf{y_k} f_c(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} \\
&E(\mathbf{y_k}|\mathcal{T}_{low,2} < \mathbf{y_k^*} < \mathcal{T}_{high,2}) = M(\mathcal{T}_{low,2}, \mathcal{T}_{high,2}) \\
&\vdots \\
&E(\mathbf{y_k}|\mathbf{y_k^*} < \mathcal{T}_{low,N}) = \int_{-\infty}^{\mathcal{T}_{low,N}} \mathbf{y_k} f_c(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}) d\mathbf{y_k}
\end{aligned}
$$

Using the above results, we can get the $E(\mathbf{y_k}|\mathbf{x_{k|k-1}})$

$$
\begin{aligned}
E(\mathbf{y_k}|\mathbf{x_{k|k-1}}) \quad = \quad & F(\mathbf{y_k} > \mathcal{T}_{high,1}) E(\mathbf{y_k}|\mathbf{y_k} > \mathcal{T}_{high,1}) \\
& + \Big(F(\mathcal{T}_{low,1} < \mathbf{y_k} < \mathcal{T}_{high,1}) \\
& \quad E(\mathbf{y_k}|\mathcal{T}_{low,1} < \mathbf{y_k} < \mathcal{T}_{high,1}) \Big) \\
& + \Big(F(\mathcal{T}_{high,2} < \mathbf{y_k} < \mathcal{T}_{low,1}) \\
& \quad E(\mathbf{y_k}|\mathcal{T}_{high,2} < \mathbf{y_k} < \mathcal{T}_{low,1}) \Big) \\
& + \Big(F(\mathcal{T}_{low,2} < \mathbf{y_k} < \mathcal{T}_{high,2}) \\
& \quad E(\mathbf{y_k}|\mathcal{T}_{low,2} < \mathbf{y_k} < \mathcal{T}_{high,2}) \Big) \\
& + \ldots + \Big(F(\mathbf{y_k} < \mathcal{T}_{low,N}) \\
& \quad E(\mathbf{y_k}|\mathbf{y_k} < \mathcal{T}_{low,N}) \Big)
\end{aligned}
$$

where,

$$
F(\alpha < \mathbf{y_k} < \beta) = \int_{\alpha}^{\beta} f(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}) d\mathbf{y_k}
$$

is the probability that $\mathbf{y_k}$ is between α and β. In simpler notation the $\mathbf{E}(\mathbf{y_k}|\mathbf{x_{k|k-1}})$ can be written as,

$$
\begin{aligned}
E(\mathbf{y_k}) \quad &= E(\mathbf{y_k^*}) - \\
&\textstyle\sum_{\eta=0}^{N} F(\mathcal{T}_{low,\eta} < \mathbf{y_k} < \mathcal{T}_{high,\eta}) E(\mathbf{y_k}|\mathcal{T}_{low,\eta} < \mathbf{y_k} < \mathcal{T}_{high,\eta})
\end{aligned}
$$

Where $E(\mathbf{y_k^*})$ is the expected value of $\mathbf{y_k}$ when there are no occlusions and $E(\mathbf{y_k^*}) = \mathbf{Cx_{k|k-1}}$ is the expected measurement for the standard Kalman filter innovation. In the case of Gaussian noise distribution the total probability that $\mathbf{y_k^*}$ is in a certain region is,

$$F(A < \mathbf{y_k^*} < B) = \int_A^B \frac{1}{\sqrt{|\mathbf{\Pi_k}|(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{y_k^*} - \mathbf{E(y_k^*)}\mathbf{\Pi_k^{-1}}(\mathbf{y_k^*} - \mathbf{E(y_k^*)}} d\mathbf{y_k^*}$$

Using the property that $F(A < \mathbf{y_k^*} < B) = F(\mathbf{y_k^*} < B) - F(\mathbf{y_k^*} < A)$. We will be able to use this property along with Bayes formula to compute $E(\mathbf{y_k})$ and $\mathbf{R_k}$

The computation for the covariance $\mathbf{R_k}$ is derived in a similar fashion as $E(\mathbf{y_k}|\mathbf{x_{k|k-1}})$. We will assume that there is no cross correlation amongst measurements, meaning $\mathbf{R_k}$ is diagonal. Equation 5.6 is the equation for $\mathbf{R_k}$, where conditions on $\mathbf{x_{k|k-1}}$ are assumed as in, $E(\mathbf{y_k}) = E(\mathbf{y_k}|\mathbf{x_{k|k-1}})$ to simplify the notation in the next sections.

$$\begin{aligned} \mathbf{R_k} &= E((\mathbf{y_k} - E(\mathbf{y_k}))(\mathbf{y_k} - E(\mathbf{y_k}))^T)) \\ &= E(\mathbf{y_k}\mathbf{y_k^T}) - E(\mathbf{y_k})E(\mathbf{y_k})^T \end{aligned} \quad (5.6)$$

We will solve for $\mathbf{R_k}$ in general,

$$\begin{aligned} E(\mathbf{y_k}\mathbf{y_k}^T) &= \int_{-\infty}^{\infty} \mathbf{y_k}\mathbf{y_k^T} f(\mathbf{y_k}|\mathbf{\Pi_k}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} \\ &= F(\mathbf{y_k^*} > \mathcal{T}_{high,1}) E(\mathbf{y_k}\mathbf{y_k}^T | \mathbf{y_k^*} > \mathcal{T}_{high,1}) \\ &\quad + \Big(F(\mathcal{T}_{low,1} < \mathbf{y_k^*} < \mathcal{T}_{high,1}) \\ &\quad\ E(\mathbf{y_k}\mathbf{y_k}^T | \mathcal{T}_{low,1} < \mathbf{y_k^*} < \mathcal{T}_{high,1}) \Big) \\ &\quad + \Big(F(\mathcal{T}_{high,2} < \mathbf{y_k^*} < \mathcal{T}_{low,1}) \\ &\quad\ E(\mathbf{y_k}\mathbf{y_k}^T | \mathcal{T}_{high,2} < \mathbf{y_k^*} < \mathcal{T}_{low,1}) \Big) \\ &\quad + \Big(F(\mathcal{T}_{low,2} < \mathbf{y_k^*} < \mathcal{T}_{high,2}) \\ &\quad\ E(\mathbf{y_k}\mathbf{y_k}^T | \mathcal{T}_{low,2} < \mathbf{y_k^*} < \mathcal{T}_{high,2}) \Big) \\ &\quad + \ldots + \Big(F(\mathbf{y_k^*} < \mathcal{T}_{low,N}) \\ &\quad\ E(\mathbf{y_k}\mathbf{y_k}^T | \mathbf{y_k^*} < \mathcal{T}_{low,N}) \Big) \\ &= E(\mathbf{y_k^*}\mathbf{y_k^*}^T) - \\ &\quad \textstyle\sum_{\eta=0}^{N} \Big(F(\mathcal{T}_{low,\eta} < \mathbf{y_k} < \mathcal{T}_{high,\eta}) \\ &\quad E(\mathbf{y_k}\mathbf{y_k}^T | \mathcal{T}_{low,\eta} < \mathbf{y_k} < \mathcal{T}_{high,\eta}) \Big) \end{aligned} \tag{5.7}$$

With $E(\mathbf{y_k}\mathbf{y_k}^T | \mathcal{T}_{low,1} < \mathbf{y_k^*} < \mathcal{T}_{high,1}) = M(\mathcal{T}_{low,1}, \mathcal{T}_{high,1})^2$ and $E(\mathbf{y_k^*}\mathbf{y_k^*}^T) = \mathbf{R}$, the variance without occlusion censoring. To complete $\mathbf{R_k}$, $E(\mathbf{y_k})E(\mathbf{y_k})^T$ should be subtracted from Equation 5.7.

The above values for $E(\mathbf{y_k})$ and $E(\mathbf{y_k}\mathbf{y_k}^T)$ can be calculated using the below equations for Gaussian integrals, for each element of array $E(\mathbf{y_k})$, and the diagonal elements of $E(\mathbf{y_k}\mathbf{y_k}^T)$. The cumulative density function and the probability density function for a normal distribution is with zero mean and unity variance,

$$\Phi(\alpha) = \int_{-\infty}^{\alpha} \phi(\alpha') d\alpha'$$

$$\phi(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}}$$

where α is a scalar. For the mean, where $\mathbf{y_k}(l)$, $\mathcal{T}_{low,\eta}(l)$, $\mathcal{T}_{high,\eta}(l)$ represent the measurements, and the lower and higher threshold limits respectively for the lth measurement.

So,

$$\begin{aligned} E(\mathbf{y_k}) &= \mathbf{Cx_{k|k-1}} \\ &\quad - \textstyle\sum_{\eta=1}^{N} \mathbf{H}(\mathbf{x}, \mathbf{Cx_{k|k-1}}, \mathbf{Diag}(\mathbf{R}), \mathcal{T}_{low,\eta}, \mathcal{T}_{high,\eta}) \end{aligned}$$

So for $\mathbf{R_k}$,

$$\begin{aligned} E(\mathbf{y_k y_k^T}) &= \mathbf{V}(\mathbf{x}, \mathbf{Cx_{k|k-1}}, \mathbf{Diag}(\mathbf{R}), -\infty, \infty) \\ &\quad - \textstyle\sum_{\eta=1}^{N} \mathbf{V}(\mathbf{x}, \mathbf{Cx_{k|k-1}}, \mathbf{Diag}(\mathbf{R}), \mathcal{T}_{low,\eta}, \mathcal{T}_{high,\eta}) \end{aligned}$$

Where $\mathbf{H}(x, \mu, \sigma, a, b)$ and $\mathbf{V}(\mathbf{x}, \mu, \sigma, \mathbf{a}, \mathbf{b})$ are defined in 4.12 and 4.11.

5.5 Problem Formulation for 2D

Using the same approach above we will derive the Tobit Kalman fitler for 2D spatial tracking applications. The difference between spatial tracking with 2D occlusion and 1D occlusions censoring is that there is a correlation between the horizontal and vertical directions uncertainty, meaning $\mathbf{R_k}$ is no longer diagonal, and assumption 1 from [53] is invalid. For the formulation of the 2D tracker in an occluded environment we will again find the censored probability distribution of the measurements, then the expected values of the measurents and the dynamic measurment noise covariance matrix. The standard state space model will be used,

$$\begin{aligned} \mathbf{x_k} &= \mathbf{Ax_{k-1}} + \mathbf{Bu_{k-1}} + \mathbf{v_k} \\ \mathbf{y_k^*} &= \mathbf{Cx_k} + \mathbf{w_k} \\ \mathbf{y_k} &= \begin{cases} \mathbf{y_k^*}, & \mathbf{Cx_k} + \mathbf{w_k} \in \mathbf{S} \\ \mathbf{T}, & otherwise \end{cases} \end{aligned} \tag{5.8}$$

where $\mathbf{S}$ is the uncensored region. For this paper we restrict ourselves to a constant velocity model for the targets which has been used previously for tracking applications [74]. The states are, $\mathbf{x_k} = [x_{pos}\ x_{pos,vel}\ y_{pos}\ y_{posvel}]$ where x_{pos} is the location of the target in the horizontal axis while $x_{pos,vel}$ is the horizontal velocity, $x_{pos},\ x_{pos,vel} \in \mathbb{R}_x$, where $\mathbb{R}_x$ is the set of real numbers on the horizontal axis. The states y_{pos} is the location of the target in the vertical axis while $y_{pos,vel}$ is the vertical velocity $y_{pos},\ y_{pos,vel} \in \mathbb{R}_y$, where $\mathbb{R}_y$ is the set of real numbers on the vertical axis . The $\mathbf{T}$ is the thresholded value, (the measurement) when the target is in an occluded region. The state space matrix's with $\Delta t = 1$,

$$\mathbf{A} = \alpha \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

The $\mathbf{v_k}$ and $\mathbf{w_k}$ in Equation 5.8 are white gaussian noise with covariance $\mathbf{Q}$ and $\mathbf{R}$ respectively.

The measurements of $\mathbf{Cx_k} + \mathbf{w_k}$ only occur when $\mathbf{Cx_k} + \mathbf{w_k} \in \mathbf{S}$, see Figure 5.3. The occluded regions will be represented by $\mathbf{O_n}$ and the region ouside the frame will be represented by $\mathbf{F}$. In this paper we will restrict our censoring to just occlusion censoring, that is, $\mathbf{F} = \mathbf{S}$.

5.6 Introduction to the 2D Gaussian Probability Density Function

Because the measurements are not deterministic we will present the statistical framework for a 2D sensor system. The second order Gaussian probability distribution will be used to describe the noise in the measurements of the 2D tracking system. For a second order measurement model we have,

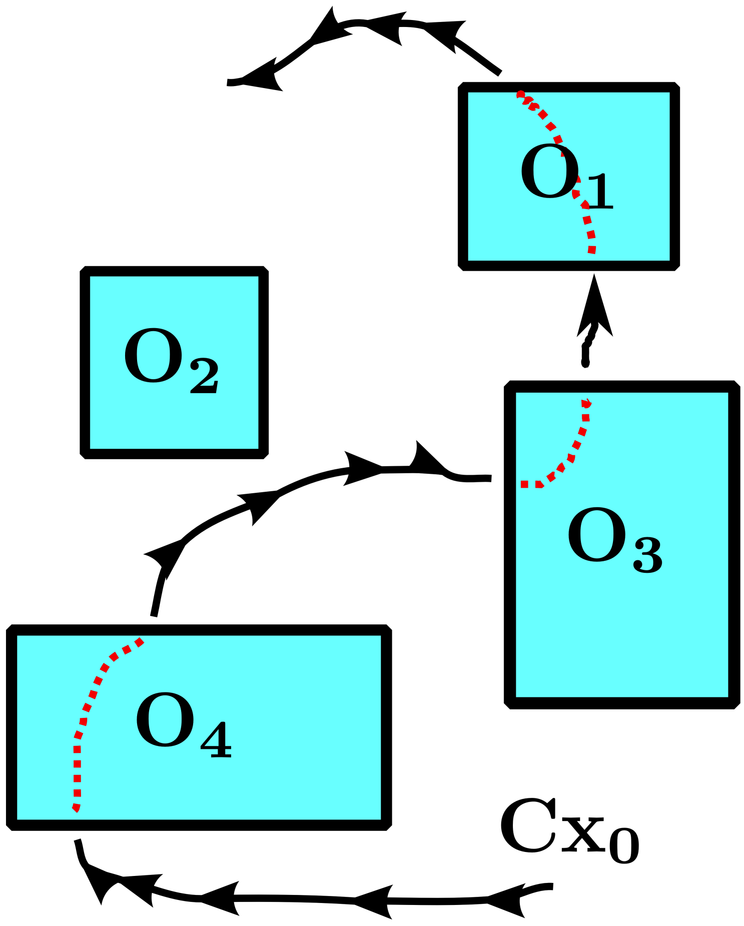

Figure 5.2: Spatial tracking space, Cx_0 is the initial measurement location

$$f(x_k, y_k) = \frac{1}{2\pi\sqrt{det(\Sigma)}} e^{-\frac{1}{2}[x_k-\mu_x \; y-\mu_y]\Sigma^-1[x_k-\mu_x \; y-\mu_y]}$$

Where x_k and y_k and locations in the horizontal and vertical directions Expanded,

$$f(x_k, y_k) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho)}\left(\frac{(x_k-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y_k-\mu_y)^2}{\sigma_y^2}\right)}$$

Where,

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

$$E(x_k) = \mu_x$$
$$E(y_k) = \mu_y$$

And ρ is the correlation coefficient,

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

For our purposes we set the correlation coefficient to $\rho = 0$ which results in independence of the x and y variables, leaving us with the following property.

$$f(x_k, y_k) = f(x_k)f(y_k)$$

The position measurements are also orthogonal since,

$$E(x_k, y_k) = E(x_k)E(y_k)$$

These results are used mostly in spatial tracking applications, however, when occlusion censoring occurs there is a cross correlation term so the results above will not fully represent the censored measurement.

5.7 Statistics for the 2D Tobit Kalman filter

First, we will define the measurement space, the occluded space and the unoccluded space. The occluded space $\mathbf{O}_\eta$ is defined as,

$$\mathbf{O}_\eta = \{(x_k, y_k) \in \mathbb{R}^2 : (O_{\eta\, x_{min}} < x_k < O_{\eta\, x_{max}}) \cap (O_{\eta\, y_{min}} < y_k < O_{\eta\, y_{max}})\}$$

And the measurement space,

$$\mathbb{M} = \{(x_k, y_k) \in \mathbb{R}^2\}$$

The space, **S**, is the unoccluded region where measurements are not censored,

$$\mathbf{S} = \mathbb{M} - \textstyle\sum_{\eta=1}^{N} \mathbf{O}_\eta$$

To model the noise on the measurements of a non censored system, when **S** covers the entire space, we will use the 2D multivariate Gaussian. **R**, is diagonal and represents the measurement error covariance matrix when there are no occlusions.

$$\mathbf{y_k} \sim \mathcal{N}(\mu, \mathbf{R})$$
$$\mathbf{R} = diag(\sigma_x, \sigma_y)$$

with,

$$f(x_k, y_k) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x_k-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y_k-\mu_y}{\sigma_y}\right)^2} \tag{5.9}$$

The μ_x and μ_y are the mean of the measurements in the horizontal and vertical directions, for the model in Equation 5.8 the mean measurements are,

$$\mu_k^x = E(\mathbf{Cx_{k|k-1}}(1))$$

$$\mu_k^y = E(\mathbf{Cx_{k|k-1}}(3))$$

Where $\mathbf{Cx_{k|k-1}}(1)$ and $\mathbf{Cx_{k|k-1}}(3)$ denotes the first and third element of the array.

As in the previous example, we will use the value of the total uncertainty using the Tobit Kalman estimate of the error covariance. To aid in calculating the probability we can use the state estimate covariance to improve the estimate of the probability that the state is uncensored.

So the uncensored measurement variance, σ_x and σ_y, will be added to the state error covariance matrix in the x and y directions. We will denote the final measurement variances for the x and y direction as Σ_x and Σ_y,

$$\Sigma_x = \sqrt{\sigma_x^2 + u^T C \Psi_{k|k-1} C^T u}$$
$$\Sigma_y = \sqrt{\sigma_y^2 + q^T C \Psi_{k|k-1} C^T q}$$

Where $u = [1\ 0]^T$ and $q = [0\ 1]^T$ are the unit directional vectors, and $u^T C \Psi_{k|k-1} C^T u$ and $q^T C \Psi_{k|k-1} C^T q$ are the respective *a priori* measurement error covariances in those directions. The covariance from the *a priori* state error covariance Σ_x and Σ_y should be replaced in the Equation 5.9, the new probability distribution is,

$$\mathbf{y} \sim \mathcal{N}(\mu, \mathbf{\Pi_k})$$
$$\mathbf{\Pi_k} = diag(\Sigma_x, \Sigma_y)$$
$$f(x_k, y_k) = \frac{1}{2\pi \Sigma_x \Sigma_y} e^{-\frac{1}{2}\left(\frac{x_k - \mu_k^x}{\Sigma_x}\right)^2 + \left(\frac{y_k - \mu_k^y}{\Sigma_y}\right)^2}$$

Where $\mathbf{\Pi_k}$ is measurement uncertainty.

This formulation is beneficial in cases where a target is missing for may subsequent measurements, the state estimate covariance grows when measurements are censored, so this value should be included, along with the measurement noise, when calculating the probability of censoring.

Where N occlusions are represented by $\mathbf{O_n}$, the probability distribution of the measurement noise is,

$$f_c(x_k, y_k) = \begin{cases} f(x_k, y_k), & \mathbf{y_k^*} \in \mathbf{S} \\ \Delta_k \int_{(x_k, y_k) \in \mathbf{O_1}} f(x_k, y_k) dx_k dy_k & \mathbf{y_k^*} \in \mathbf{O_1} \\ \vdots \\ \Delta_k \int_{(x_k, y_k) \in \mathbf{O_N}} f(x_k, y_k) dx_k dy_k, & \mathbf{y_k^*} \in \mathbf{O_N} \end{cases}$$

with $\Delta_k = \delta(x_k - \tau_x, y_k - \tau_y)$, the Kronecker delta function in two dimensions. See Figure 5.3 for a sample distribution of measurement uncertainty with one occlusion.

The expected value of the measurements is given,

$$
\begin{aligned}
E(\mathbf{y_k}|\mathbf{y_k^*} \in \mathbf{S}) &= \int_{(x_k,y_k)\in\mathbf{S}} \mathbf{y_k} f(x_k, y_k) dx_k dy_k \\
E(\mathbf{y_k}|\mathbf{y_k^*} \in \mathbf{O_1}) &= [\mathcal{T}_x\ \mathcal{T}_y]^T_{\mathbf{O_1}} \\
E(\mathbf{y_k}|\mathbf{y_k^*} \in \mathbf{O_2}) &= [\mathcal{T}_x\ \mathcal{T}_y]^T_{\mathbf{O_2}} \\
&\vdots \\
E(\mathbf{y_k}|\mathbf{y_k^*} \in \mathbf{O_N}) &= [\mathcal{T}_x\ \mathcal{T}_y]^T_{\mathbf{O_N}}
\end{aligned}
$$

In occluded regions $[\mathcal{T}_x\ \mathcal{T}_y]^T_{\mathbf{O_1}} = COM(\mathbf{O_1})$, and $COM(\mathbf{O_1}) \in \mathbb{R}^{2\times1}$ is the center of mass of $\mathbf{O_1}$ assuming uniform "density."

$$
COM(\mathbf{O_1}) = \begin{bmatrix} \frac{1}{\mathbf{O_1}_{\,\mathbf{x_{max}}} - \mathbf{O_1}_{\,\mathbf{x_{min}}}} \int_{\mathbf{O_1}_{\,\mathbf{x_{min}}}}^{\mathbf{O_1}_{\,\mathbf{x_{max}}}} \xi d\xi \\ \frac{1}{\mathbf{O_1}_{\,\mathbf{y_{max}}} - \mathbf{O_1}_{\,\mathbf{y_{min}}}} \int_{\mathbf{O_1}_{\,\mathbf{y_{min}}}}^{\mathbf{O_1}_{\,\mathbf{y_{max}}}} \nu d\nu \end{bmatrix}
$$

Where $\mathbf{O_1}_{\,\mathbf{x_{min}}}, \mathbf{O_1}_{\,\mathbf{x_{max}}}$ and $\mathbf{O_1}_{\,\mathbf{y_{min}}}, \mathbf{O_1}_{\,\mathbf{y_{max}}}$ are the maximum and minimum values in the horizontal and vertical axis of $\mathbf{O_1}$.

5.7.1 The Statistics for a 2D Tobit Kalman Filter

In this section $\mathbf{E}(\mathbf{y_k})$ and $\mathbf{R_k}$ with N occlusions is calculated. To compute the statistics for the 2D Tobit Kalman Filter we assume that the occlusions are rectangular to allow numerical solutions for the mean and variance equations. The distribution will be centered at the model mean, which is $\mathbf{Cx_{k|k-1}}$ and measurement covariance given by the variance of the measurement noise. The probability of the measurement being in the nth occlusion is,

$$
F(x_k, y_k | x_k, y_k \in O_n) = \int_{O_1\ x_{min}}^{O_1\ x_{max}} \int_{O_1\ y_{min}}^{O_1\ y_{max}} f(x_k, y_k) dx_k dy_k
$$

The probability of being uncensored is,

$$
F(x_k, y_k | x_k, y_k \in \mathbf{S}) = 1 - \sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)
$$

To continue with the statistics in the 2D space, we will introduce the notation for image moments. The moment function is denoted by,

$$\mathbf{M}_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u^p v^q f(u,v) du dv$$

the general form for a central moment is,

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u - \bar{u})^p (v - \bar{v})^q f(u,v) du dv$$

Where $\bar{u} = \mathbf{M_{10}}$ and $\bar{v} = \mathbf{M_{01}}$.

To calculate the expected value of the measurements using the model in Equation 5.8, we will use the notation for image moments. The mean in the x direction and the mean in y direction will be represented as $[\mathbf{M}_{10}\ \mathbf{M}_{01}]^T$ when there is no censoring. We use the notation, $\mathbf{M}_{10|S}$ to represent the first moment in the x direction, in the space $\mathbf{S}$ and $\mathbf{M}_{10|\mathbf{O}_\eta}$ to represent the first moment in the x direction, in the occlusion $\mathbf{O}_\eta$. To calculate the mean for the Tobit Kalman filter we continue with Equation 5.7, and calculate $\mathbf{y_k} \in S$,

$$\begin{aligned} \mathbf{M}_{10|\mathbf{S}} &= \textstyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_k f(x_k, y_k) dx_k dy_k \\ &- \textstyle\sum_{\eta=1}^{N} \int_{O_\eta\, x_{min}}^{O_\eta\, x_{max}} \int_{O_\eta\, y_{min}}^{O_\eta\, y_{max}} x_k f(x_k, y_k) dx_k dy_k \end{aligned}$$

where the image moment in occlusionη is,

$$\mathbf{M}_{10|\mathbf{O}_\eta} = \textstyle\int_{O_\eta\, x_{min}}^{O_\eta\, x_{max}} \int_{O_\eta\, y_{min}}^{O_\eta\, y_{max}} x_k f(x_k, y_k) dx_k dy_k$$

$\mathbf{M}_{10|\mathbf{S}}$, $\mathbf{M}_{10|\mathbf{O}_\eta}$ represents the first moment in the horizontal direction in the space $\mathbf{S}$, and behind an occlusion, η. The moment in the vertical direction is,

$$\begin{aligned} \mathbf{M}_{01|\mathbf{S}} &= \textstyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_k f(x_k, y_k) dx_k dy_k \\ &- \textstyle\sum_{\eta=1}^{N} \int_{O_\eta\, x_{min}}^{O_\eta\, x_{max}} \int_{O_\eta\, y_{min}}^{O_\eta\, y_{max}} y_k f(x_k, y_k) dx_k dy_k \end{aligned}$$

where,

$$\mathbf{M}_{01|\mathbf{O_n}} = \textstyle\int_{O_n\, x_{min}}^{O_n\, x_{max}} \int_{O_n\, y_{min}}^{O_n\, y_{max}} y_k f(x_k, y_k) dx_k dy_k \qquad (5.10)$$

The total model mean for the entire space is equal to,

$$\mathbf{M}_{01|\mathbf{S}} = \mathbf{M}_{01} - \sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)\mathbf{M}_{01|\mathbf{O_n}} \tag{5.11}$$

For the x_k direction,

$$\mathbf{M}_{10|\mathbf{S}} = \mathbf{M}_{10} - \sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)\mathbf{M}_{10|\mathbf{O_n}} \tag{5.12}$$

Where $\mathbf{M}_{01}$ and $\mathbf{M}_{01}$ represent moments in an uncensored situation, μ_k^x and μ_k^y respectively. Because we are assuming that the occlusions are rectangular we can solve for the elements in $\mathbf{O_n}$.

$$F(x_k, y_k | x_k, y_k \in \mathbf{O_n}) = (\Phi(\tfrac{O_{n\ xmax} - \mu_k^x}{\sigma}) - \Phi(\tfrac{O_{n\ xmin} - \mu_k^x}{\sigma}))$$
$$(\Phi(\tfrac{O_{n\ ymax} - \mu_k^y}{\sigma}) - \Phi(\tfrac{O_{n\ ymin} - \mu_k^y}{\sigma}))$$

So the moments for Equations 5.11 and 5.12 starting with Equation 5.10 is,

$$\mathbf{M}_{10|\mathbf{O_n}} = COM(\mathbf{O_n})(1)$$

and in the y direction,

$$\mathbf{M}_{01|\mathbf{O_n}} = COM(\mathbf{O_n})(2)$$

where $COM(\mathbf{O_n})(1)$ is the first element in $COM(\mathbf{O_n})$. The above moment is the expected value of censored data which is a constant, so they are simple to compute. In censored conditions, where a measurement lies behind an occlusion, the expected measurement is an arbitrary value we choose to be the center of mass of an occlusion.

Using $\mathbf{H}$ in Equation 4.12 from the single dimension case, the first moment in the x direction is,

$$\mathbf{M}_{10|\mathbf{S}} = \mathbf{M}_{10} - \sum_{n=1}^{N} \left(\mathbf{H}(\mathbf{x}, \mu, \sigma, \mathbf{O}_{\mathbf{n}\ \mathbf{x_{max}}}, \mathbf{O}_{\mathbf{n}\ \mathbf{x_{min}}} \left(\Phi\left(\frac{O_{n\ y_{max}} - \mu_y^k}{\sigma}\right) - \Phi\left(\frac{O_{n\ y_{min}} - \mu_y^k}{\sigma}\right)\right) \right)$$

The first moment in the y direction is,

$$\mathbf{M}_{01|\mathbf{S}} = \mathbf{M}_{10} - \sum_{n=1}^{N} \left(\mathbf{H}(\mathbf{y}, \mu, \sigma, \mathbf{O}_{\mathbf{n}\ \mathbf{y_{max}}}, \mathbf{O}_{\mathbf{n}\ \mathbf{y_{min}}} \left(\Phi\left(\frac{O_{n\ x_{max}} - \mu_x^k}{\sigma}\right) - \Phi\left(\frac{O_{n\ x_{min}} - \mu_x^k}{\sigma}\right)\right) \right)$$

The above equations are computing the moments by finding the total expectation in the space then subtracting segments contained in occluded areas. To continue as in Section 5.4 for the 1D case, the next step for the Tobit Kalman filter is to find the values of the dynamic measurement covariance matrix, $\mathbf{R_k}$. We will continue to use the image moment notation, so finding the second moments in $\mathbf{S}$ in the x direction is,

$$\mathbf{M}_{20|\mathbf{S}} = \mathbf{M}_{20} - \sum_{n=1}^{N} \left(\mathbf{V}(\mathbf{x}, \mu, \sigma, \mathbf{O}_{\mathbf{n}\ \mathbf{x_{max}}}, \mathbf{O}_{\mathbf{n}\ \mathbf{x_{min}}} \left(\Phi\left(\frac{O_{n\ y_{max}} - \mu_y^k}{\sigma}\right) - \Phi\left(\frac{O_{n\ y_{min}} - \mu_y^k}{\sigma}\right)\right) \right) \tag{5.13}$$

The second moment in the y direction is,

$$\mathbf{M}_{02|\mathbf{S}} = \mathbf{M}_{02} - \sum_{n=1}^{N} \left(\mathbf{V}(\mathbf{y}, \mu, \sigma, \mathbf{O}_{\mathbf{n}\ \mathbf{y_{max}}}, \mathbf{O}_{\mathbf{n}\ \mathbf{y_{min}}} \left(\Phi\left(\frac{O_{n\ x_{max}} - \mu_x^k}{\sigma}\right) - \Phi\left(\frac{O_{n\ x_{min}} - \mu_x^k}{\sigma}\right)\right) \right) \tag{5.14}$$

The nature of the example in this paper is that there is a heavy dependence between the x and y states. This is apparent in the cross terms of the measurement noise matrix $\mathbf{R_k}$. The cross dependence between x and y is apparent in the second moment $\mathbf{M}_{11|\mathbf{S}}$ when there are censored measurements.

$$\begin{aligned}
&\mathbf{M}_{11|\mathbf{S}} = \mathbf{M}_{11} \\
&-\textstyle\sum_{n=1}^{N} \Big(\mathbf{H}(\mathbf{x}, \mu_{\mathbf{x}}, \sigma, \mathbf{O_{n\,x_{max}}}, \mathbf{O_{n\,x_{min}}}) \\
&\quad \mathbf{H}(\mathbf{y}, \mu_{\mathbf{y}}, \sigma, \mathbf{O_{n\,y_{max}}}, \mathbf{O_{n\,y_{min}}})\Big)
\end{aligned} \tag{5.15}$$

The above Equations 5.13-5.15 are only applicable when there is no cross dependence in the horizontal and vertical directions of the latent measurement noise, $\mathbf{R}$ is diagonal. The complete calculation for the moments $\mathbf{M}_{10|\mathbb{M}}$, $\mathbf{M}_{01|\mathbb{M}}$,$\mathbf{M}_{20|\mathbb{M}}$, $\mathbf{M}_{02|\mathbb{M}}$, and $\mathbf{M}_{11|\mathbb{M}}$ of the entire censored measurement space M is,

$$\begin{aligned}
\mathbf{M}_{10|\mathbb{M}} &= F(x_k, y_k | x_k, y_k \in S)\mathbf{M}_{10|\mathbf{S}} + \\
&\textstyle\sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)\mathbf{M}_{10|\mathbf{O_n}}
\end{aligned}$$

$$\begin{aligned}
\mathbf{M}_{01|\mathbb{M}} &= F(x_k, y_k | x_k, y_k \in S)\mathbf{M}_{01|\mathbf{S}} + \\
&\textstyle\sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)\mathbf{M}_{01|\mathbf{O_n}}
\end{aligned}$$

$$\begin{aligned}
\mathbf{M}_{20|\mathbb{M}} &= F(x_k, y_k | x_k, y_k \in S)\mathbf{M}_{20|\mathbf{S}} + \\
&\textstyle\sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)\mathbf{M}_{20|\mathbf{O_n}}
\end{aligned}$$

$$\begin{aligned}
\mathbf{M}_{02|\mathbb{M}} &= F(x_k, y_k | x_k, y_k \in S)\mathbf{M}_{02|\mathbf{S}} + \\
&\textstyle\sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)\mathbf{M}_{02|\mathbf{O_n}}
\end{aligned}$$

$$\begin{aligned}
\mathbf{M}_{11|\mathbb{M}} &= F(x_k, y_k | x_k, y_k \in S)\mathbf{M}_{11|\mathbf{S}} + \\
&\textstyle\sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in O_n)\mathbf{M}_{11|\mathbf{O_n}}
\end{aligned}$$

In an occlusion, the measurements have no variance because they are a constant values known *a priori*, so,

$$\begin{aligned}
\mathbf{M}_{20|\mathbf{O_n}} &= COM(\mathbf{O_n})(1)^2 \\
\mathbf{M}_{02|\mathbf{O_n}} &= COM(\mathbf{O_n})(2)^2 \\
\mathbf{M}_{11|\mathbf{O_n}} &= COM(\mathbf{O_n})(1)COM(\mathbf{O_n})(2)
\end{aligned}$$

Where $\mathbf{0_{2x2}}$ is a zero matrix.

5.8 The 2D Tobit Kalman filter with Occluded Region Censoring

Using equations in Section 5.7.1 we can present the 2D Tobit Kalman filter with occlusion region censoring. Extending from Equations 5.2 the value of $\mathbf{E}(\mathbf{y_k})$ and $\mathbf{R_k}$ are,

$$\mathbf{E}(\mathbf{y_k}) = \begin{bmatrix} \mathbf{M}_{10} \\ \mathbf{M}_{01} \end{bmatrix}$$

$$\mathbf{R_k} = \begin{bmatrix} \mathbf{M}_{20|\mathbf{S}} - \mathbf{M}^2_{10|\mathbf{S}} & \mathbf{M}_{11|\mathbf{S}} - \mathbf{M}_{10|\mathbf{S}}\mathbf{M}_{01|\mathbf{S}} \\ \mathbf{M}_{11|\mathbf{S}} - \mathbf{M}_{10|\mathbf{S}}\mathbf{M}_{01|\mathbf{S}} & \mathbf{M}_{02|\mathbf{S}} - \mathbf{M}^2_{01|\mathbf{S}} \end{bmatrix}$$

Where $\mathbf{M}_{20|\mathbf{S}} - \mathbf{M}^2_{10|\mathbf{S}}$, $\mathbf{M}_{11|\mathbf{S}} - \mathbf{M}_{10|\mathbf{S}}\mathbf{M}_{01|\mathbf{S}}$, $\mathbf{M}_{02|\mathbf{S}} - \mathbf{M}^2_{01|\mathbf{S}}$ are derived in the previous section.

5.9 Comparison to the Standard Kalman Filter

The Kalman filter is optimal in the linear case, hence, it is important that the Tobit Kalman filter will converge to the standard Kalman filter when there are no censored regions.

5.9.1 Convergence to the Standard Kalman Filter

In the case where we have zero occlusions (N=0) then the Tobit Kalman Filter will converge to a standard Kalman filter. In [53] the Tobit Kalman filter is shown to converge to the standard Kalman filter when the probability of being uncensored goes to one, which happens in the case of zero occlusions and infinity large frame.

$$F(x_k, y_k | x_k, y_k \in \mathbf{S})$$
$$= 1 - \sum_{n=1}^{N} F(x_k, y_k | x_k, y_k \in \mathbf{O_n}) = 1$$

And the innovation process will converge to the Kalman error,

$$\begin{aligned} \lim_{(x_k,y_k)\in W} y_k - \mathbf{E}(\mathbf{y_k}) &= y_k - \begin{bmatrix} \mathbf{M}_{10|\mathbb{M}} \\ \mathbf{M}_{01|\mathbb{M}} \end{bmatrix} \\ &= y_k - Cx_{k|k-1} \end{aligned}$$

When the measurement space, $\mathbb{M} = \mathbf{S}$, the measurement covariance matrix, $\mathbf{R_k}$ will converge to,

$$\lim_{(x_k, y_k) \in \mathbf{S}} \mathbf{R_k} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

The above derivation of convergence to the standard Kalman filter can also be done by stating that the expected measurement is far away from any occlusion. The definition of 'far away' is dependent on $\mathbf{R}$. If $min(E(\mathbf{y_k}) - \mathbf{O}_\eta) >> \delta[\mathbf{R}(1,1)\ \mathbf{R}(2,2)]^T$ where δ is an scalar integer.

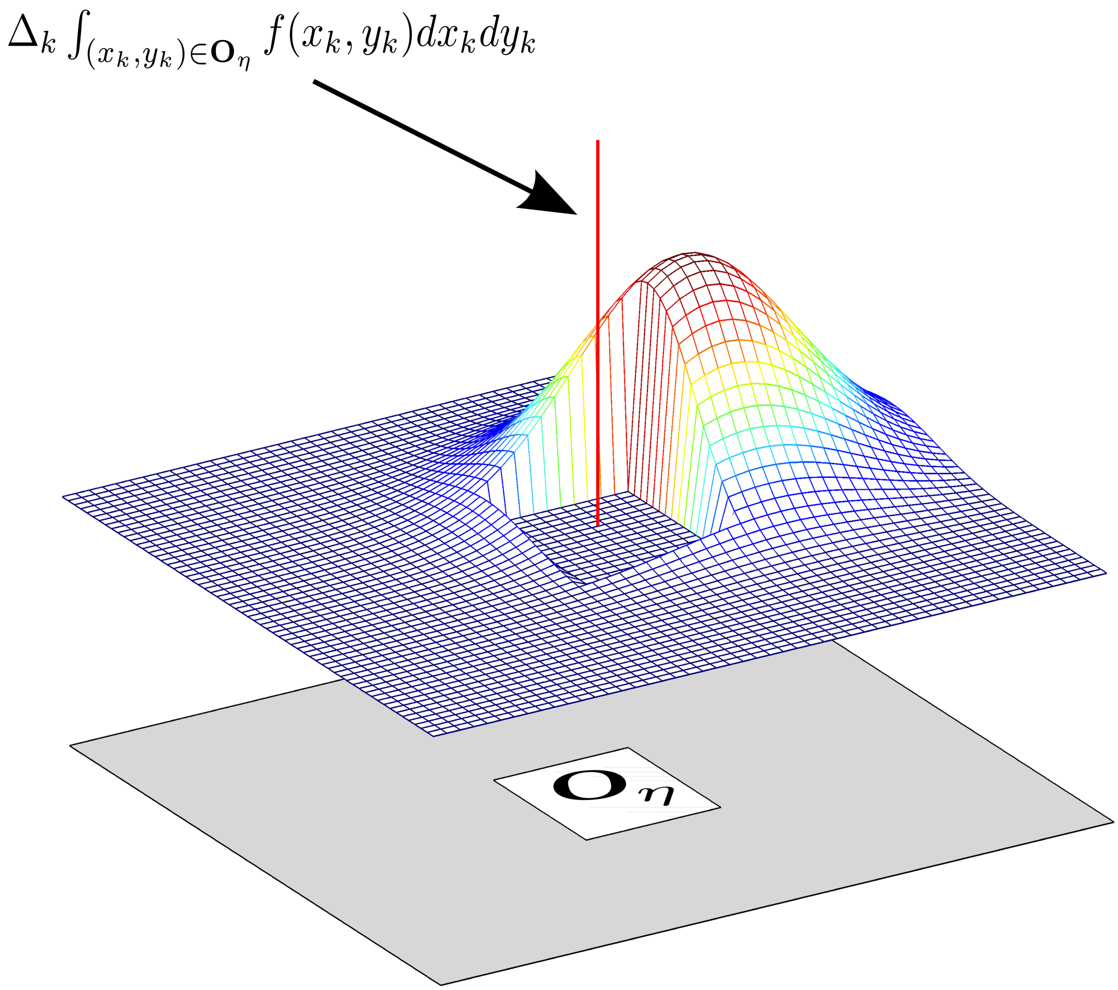

Figure 5.3: Probability distribution of 2D Gaussian with one occlusion

Chapter 6

SIMULATIONS

In this chapter, we present simulations of the Tobit Kalman filter to show usefulness in applications where censored data is unavoidable. The Tobit Kalman filter is compared to the Kalman filter for intermittent measurements (KFIM) which is outlined in [26] . This filter operates as a Kalman filter until a missing measurement occurs, then it will only predict the current state and covariance matrix and not update the filter; so $\mathbf{x_{k|k}} = \mathbf{x_{k|k-1}}$ and $\mathbf{\Psi_{k|k}} = \mathbf{\Psi_{k|k-1}}$. The KFIM treats censored measurements as missing, instead of censored. In addition to the KFIM, we compare the Tobit Kalman filter to the results of the standard Kalman filter (SKF), which treats censored values as regular measurements, and the particle filter, which is a nonlinear approach that makes no assumptions on the model or noise distributions. Many types of particle filtering are available to use [75], in this comparison we will use the sequential importance resampling (SIR), with re sampling if the effective number of particles is less than 50% of the particles used. We will use a simple re sampling strategy, the systematic re sampling approach [55]. The probability of measurements used for the particle filter weighting function, $P(y_k|x_k)$ when censoring occurs is given in Chapter 3 for 1D censoring, and Chapter 5 for 2D censoring.

6.1 One Sided Censoring

The results from three experiments demonstrate the improvement possible with the Tobit Kalman filter in comparison to the SKF, the KFIM and the particle filter. The first simulations estimate a constant value near a censoring limit and show that the Tobit Kalman Filter is unbiased. The next two simulations are a Brownian motion model and a sinusoidal motion model which have disturbances as well as additive noise.

6.1.1 Estimate a Constant Value

In this example we will estimate a constant value near a censoring region. In Figure 6.1 we have a constant value at .2 with measurement noise $\sigma = 1$ and censoring limit $\tau = 0$. The initial conditions are $x_0 = 2$ and $\Psi_0 = 1$, with $Q = 10^{-11}$. The particle filter is represented with 100 particles in this example.

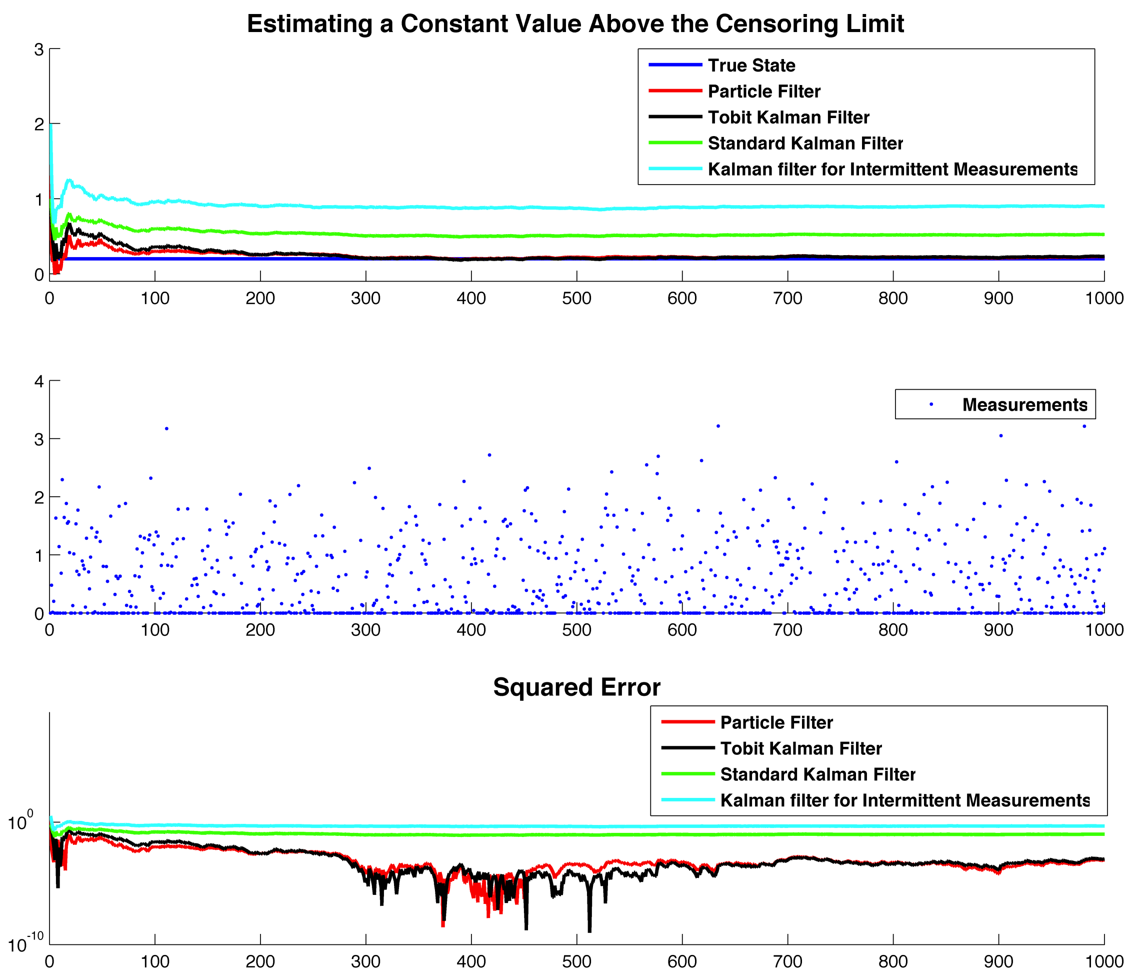

Figure 6.1: Comparison of particle filter, Tobit Kalman filter, SKF and KFIM in estimating a constant signal above the censoring limit

As shown in Figure 6.1, the Tobit Kalman filter and particle filter converge to the true value. The other methods are biased with estimates produced by the SKF and KFIM falsely converging to values in the uncensored region. In this example, the SKF converges to the average value of measurements including the censored values while the KFIM converges to the average of measurements not including the censored values.

In Figure 6.2 the particle filter and the Tobit Kalman filter are able to estimate a constant value below the censoring limit.

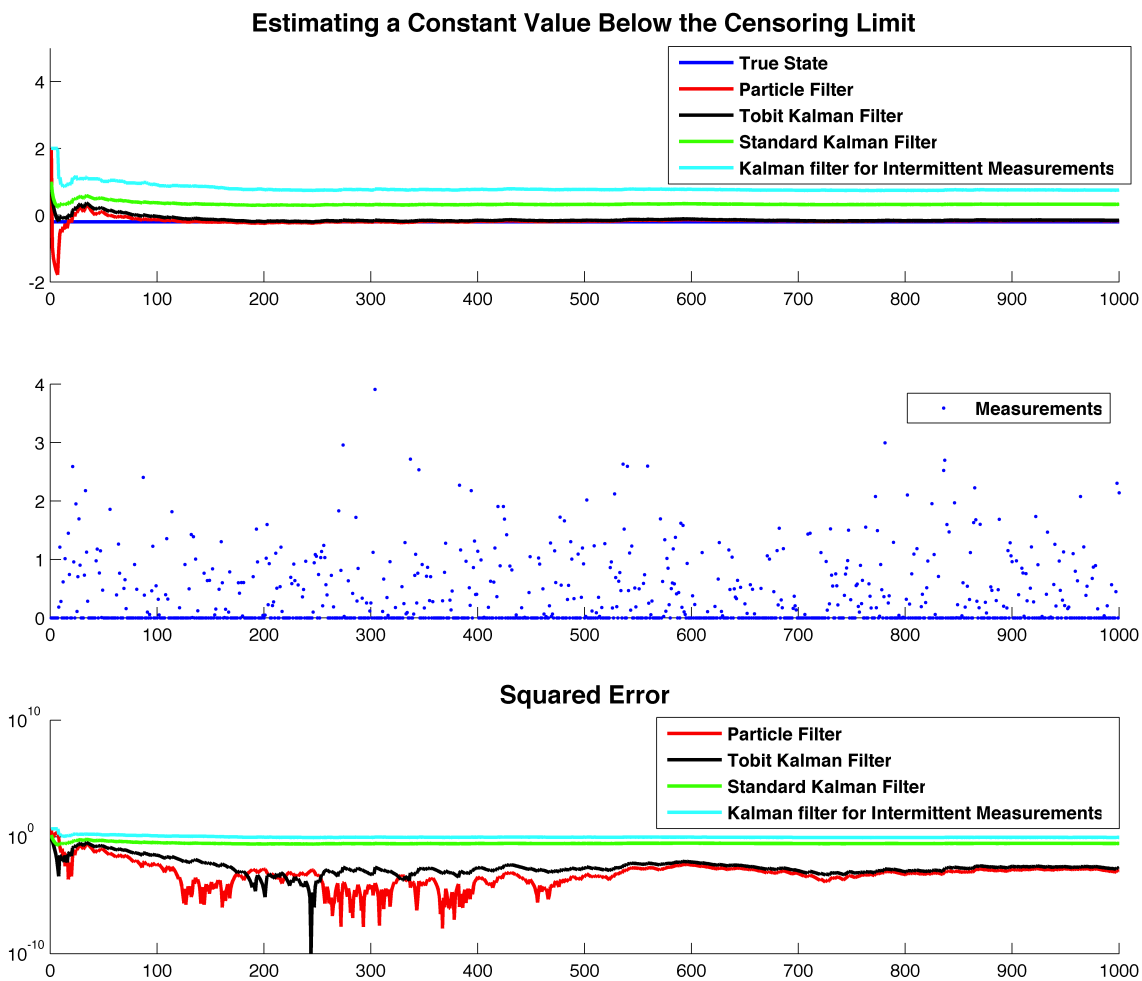

Figure 6.2: Comparison of particle filter, Tobit Kalman filter, SKF and KFIM in estimating a constant signal below the censoring limit

The previous simulations show the performance when a large number of particles is used, in Figure 6.3 the same simulation is run as in Figure 6.2, however; only 10 particles are used. In 6.3 the Tobit Kalman filter is able to outperform the particle filter, with the same conditions Figure 6.4 shows that the particle filter outperforms the Tobit Kalman filter in convergence speed but has a smaller steady state error. The variance in the steady state estimate of the Tobit Kalman filter is due to the $\mathbf{Q}$ not being zero.

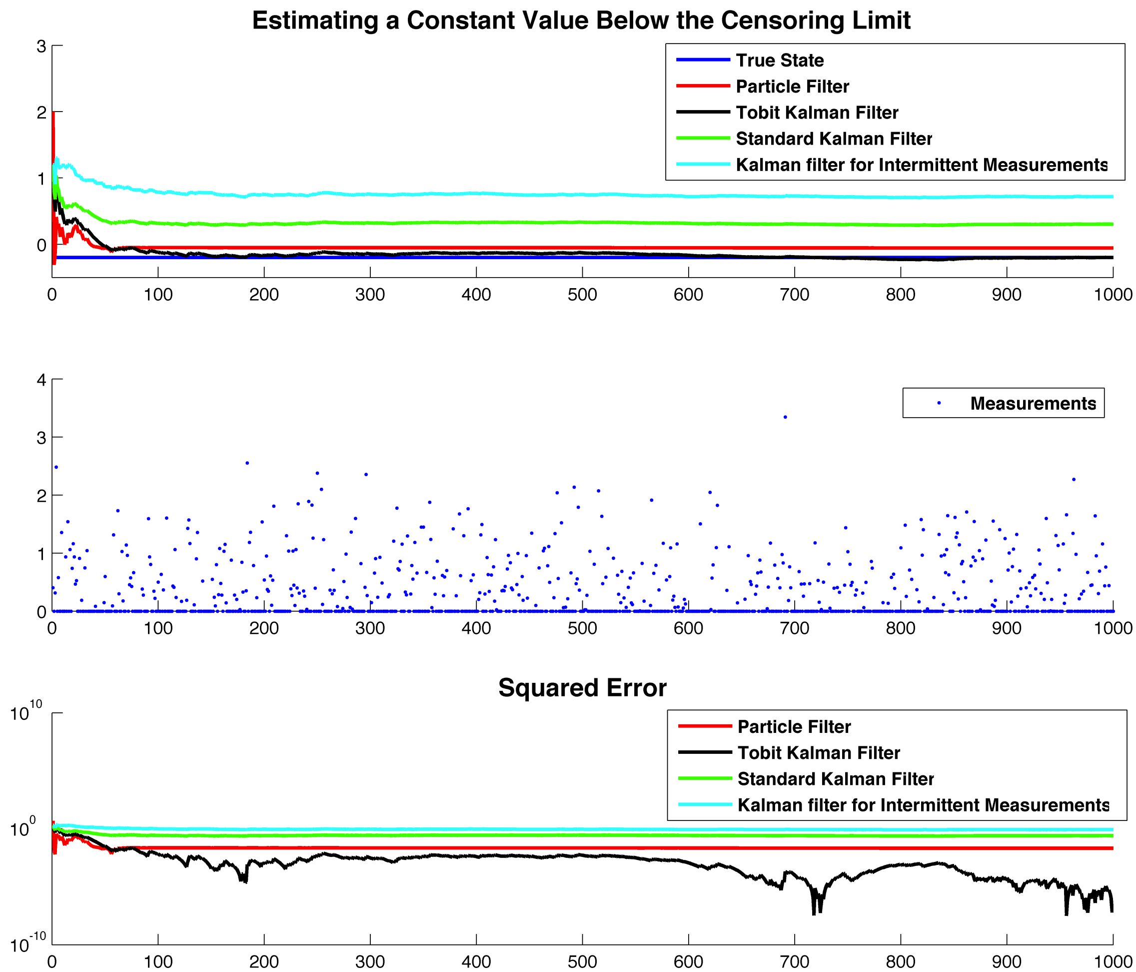

Figure 6.3: Experiment 1 with too few particles: Constant value below the censoring limit, the particle filter is run with 10 particles.

The simulations in Figure 6.3 and 6.4 show the practical effect of using a statistical approach such as a particle filter, verses a deterministic approach such as the Tobit Kalman filter. Using statistical filters can often produce undesirable results.

6.1.1.1 Random Walk Simulation

In this section, we track a damped random walk signal with normal process noise. Random walk often leads to saturation issues in both MEMS sensors and tracking occlusions with visual targets. The model is simple yet shows the tracking performance of our method in a disturbance-driven model. The data is generated from,

$$y_k^* = \alpha y_{k-1}^* + \eta_k \tag{6.1}$$

We choose $\alpha = .99$ to dampen the natural divergence of the model, keeping it close to the censoring limit. η is a normally distributed random variable with standard deviation 0.3, and the measurements have noise with $\sigma = 1$ and are left-censored at $T = 0$. The initial conditions are $\mathbf{x_0} = \mathbf{5}$ and $\mathbf{\Psi_0} = \mathbf{1}$.

As shown in Figure 6.5, the SKF's estimates converge to the censoring limit when the measurements are censored for a sufficient period. The KFIM only updates when measurements are not censored, resulting in a increased uncertainty in the state estimate as described in [26], this leads to an increased reliance on non-censored measurements. The particle filter is run with 100 particles for this example.

In this example the particle filter outperforms the Tobit Kalman filter for the censored data case, the RMS error of the particle filter is 0.6124 and for the Tobit Kalman filter, 0.6561; however, the particle filter requires a much higher computational cost.

6.1.1.2 Oscillator Simulation

The following example has state-space dynamics with the state transition and measurement transfer matrices,

$$A = \alpha \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix} \tag{6.2}$$

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

The purpose of this simulation is to show a robust tracking ability with a known model and unknown disturbance that enters the system through $\mathbf{w_k}$. In this example, $\alpha = 1$, the disturbance is normally distributed with standard deviation of .1 and is uncorrelated to the measurement noise which is normally distributed with standard deviation of $\sigma = 1$. The initial conditions are $\mathbf{x_0} = [\mathbf{5\ 0}]^{\mathbf{T}}$ and $\mathbf{\Psi_0} = \mathbf{I_{2x2}}$ and the frequency is $\omega = .007x2\pi$ with sampling period $T = 1$. This model has a relatively small process noise, so the particle filter with 100 particles has a long convergence time.

Figure 6.6 shows that even when the measurements are censored, the output of the Tobit Kalman filter closely tracks the actual state while the KFIM method and SKF are unable to track through censoring. This is because the KFIM will trust stray, non-censored data after several censored measurements; due to the large state error covariance not performing a measurement update. The SKF will converge to censored data while the state error covariance continuously decreases, even when the measurements are censored.

The particle filter comes the closest to performing as well as the Tobit Kalman filter, however, the Tobit Kalman filter far outperforms the particle filter in its convergence speed. If the process noise of the measurements increases, the particle filter converges faster, see Figure 6.7 for the same simulation but with the process noise increased to a standard deviation of .5. The particle filter is also quick to converge to measurements when the sinusoidal data re enters the uncensored region, this is direct consequence of assumption 1 of the Tobit Kalman filter which states that the probability of the measurement begin censored can be predicted by the system model. It should be noted that artificially increasing the process noise to improve convergence time will cause unnecessary large variance in the particle filter estimate, trading quicker convergence for large steady state error.

In Figure 6.8, the average RMS error for the same sinusoidal example is shown for the particle filter with 10-560 particles at 50 particle increments, where each simulation is run with 50 sets of simulated data. The process noise has standard deviation .1 and the measurement noise is $\sigma = 1$, the initial condition for the simulated data and the estimator is $x = [5\ 0]^T$. The Tobit Kalman filter outperforms the particle filter in RMS error for most cases, and has smaller variance in the performance amongst different data sets. Again, this results from the Tobit Kalman filter being deterministic while the particle filter is stochastic. This simulation shows that the stochastic approach has a significant probability of being worse than the Tobit Kalman filter even when 500 particles are used.

There are certain values for process noise where the particle filter outperforms the Tobit Kalman filter, but always at the expense of longer computation time. In Figure 6.8 the RMS error is plotted, along with time; showing that the particle filter is several times more computationally expensive than the Tobit Kalman fitler. The simulation is run with 1000 time steps and the Tobit Kalman filter will run in .3 seconds, where the particle filter will run in 2.17 seconds for 10 particles and 110.4 seconds with 560 particles.

6.2 UKF and EKF Performance

In this section we present a simulation using a sinusoidal model to compare the Tobit Kalman filter with the EKF and UKF formulation. Using the statistics described in [22] [76], the state space model we will be using is,

$$\begin{aligned} \mathbf{x_k} &= \mathbf{A}\mathbf{x_{k-1}} + \mathbf{B}\mathbf{u_k} \\ y_k^* &= \mathbf{C}\mathbf{x_k} \\ y_k &= \begin{cases} y_k^*, & y_k^* < T \\ T, & otherwise \end{cases} \end{aligned} \tag{6.3}$$

$$\mathbf{A} = \alpha \begin{bmatrix} cos(\omega) & -sin(\omega) \\ sin(\omega) & cos(\omega) \end{bmatrix} \tag{6.4}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.5}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{6.6}$$

With $\alpha < 1$

In Figure 6.9 we have a latent measurement noise of $v_k \sim \mathcal{N}(0, 2)$ and a disturbance noise of $w_k \sim \mathcal{N}(0, .1)$. The RMS error in the state estimate for the Tobit Kalman filter is less than the UKF and EKF error. The state error covariance for the UKF and EKF has a sharp discontinuity when leaving a censored region, while the Tobit Kalman filter state error covariance has a smoother transitions. This transition would be sharper in the Tobit Kalman filter if the measurement noise was smaller.

In the next example we use a less informative model, the state space changes to $\mathbf{A} = .999$, $\mathbf{B} = 1$ and $\mathbf{C} = 1$, and latent measurement noise of $v_k \sim \mathcal{N}(0, 1)$ and a disturbance noise of $w_k \sim \mathcal{N}(0, .1)$.

In Figure 6.10 the Tobit Kalman filter is able to smoothly transition from censored to non-censored regions. See the top graph of Figure 6.10 at sample 2000-2500, the Tobit Kalman filter is able to converge to the true estimate when seeing only the noise on the measurements. The EKF and UKF are not able to estimate the state unless it has exited the censored region. Applications for this type of performance in an estimator are [77], which provides an example of a stable controller when measurements are censored.

6.3 Occlusion Results

In this section we present simulations from the algorithm developed in Chapter 5. The simulation for the 1D occlusions will have the same dynamic model as in the previous section.

6.3.1 1D Occlusion Results

In the simulation depicted in Figure 6.11 there are two occluded regions with a random walk dynamical model. The process noise standard deviation for this example is 0.1 and the measurement noise is 1. The $\alpha = .999$ and the initial conditions for the state is $x_0 = 8$ and the state error covariance is $P_0 = 1$. The SKF converges faster to censored measurements in the occluded regions near samples 1140 for example, and erratic behavior when the true value is is near an occluded region. The erratic behavior near samples 1250-1280 is caused by the significant number of measurements that jump to the censoring limit, which the SKF treats as true measurements. The KFIM jumps from uncensored region to uncensored region because it only observes uncensored measurements, meanwhile the state error covariance increases with each missed measurement. The performance of the particle filter, run with 100 particles, and the Tobit Kalman filter are very similar. It is important to note that when there is a long duration of uncensored measurements both the estimates of the Tobit Kalman filter and the particle filter converge to the center of the occluded region since this is the chosen threshold value.

6.3.2 2D Occlusion Results

In this section we will present a simulation using the statistics and state space model in Chapter 3 to compare the Tobit Kalman filter to the the particle filter for spatial tracking. The particle filter is run with 100 particles in this example. In Figure 6.12 we have disturbance with standard deviation .2 so $Q = .2^2 I_{2\times 2}$, measurement noise $\sigma = 10$, $\mathbf{\Psi_0} = \mathbf{I_{4\times 4}}$the dampening factor, $\alpha = .99$, the initial conditions $\mathbf{x_k} = [\mathbf{0\ 0\ 0\ 0}]^{\mathbf{T}}$ for the trackers and simulated data. For visual purposes, the frames are organized as follows, frames 1 and 2 represent the first $1 - \frac{M}{6}$ and $1 - \frac{M}{3}$, frames 3 and 4 represent samples $\frac{M}{3} - \frac{M}{2}$ and $\frac{M}{3} - \frac{2M}{3}$, and frames 5 and 6 represent samples $\frac{2M}{3} - \frac{5M}{6}$ and $\frac{2M}{3} - M$. The Tobit Kalman filter will converge to the center of the occlusion if there is a continuous stream of occluded measurements. In the example in Figure 6.12, censored measurements do not cause the Tobit Kalman filter to drift from the actual values, see frame 5, where the state passes through the occluded region and the Tobit Kalman filter does not preemptively drift to the center of the occlusion even though there are several measurements at this point. The particle filter fails immediately and is unable to recover. This failure is due to the fact that the particle filter is a stochastic approach and has a degeneracy problem. The particle filter often fails when the motion abruptly changes, and the weighting function is not able to produce particles with high weights for the measurements received. Increasing the number of particles will decrease the probability of failure occurring. In Figure 6.13 we have the same simulation, but in this case the particle filter is able to track throughout the simulation. In frame 2, the Tobit Kalman filter is outperforming the particle filter in the occlusion, and in frames 5 and 6 the true values are changing direction several times within the occlusion with respect to the number of samples so the trackers are not performing well.

In Figure 6.14 the Tobit Kalman filter is shown to be unbiased with a constant value in a censored region. The measurement noise has $\sigma = 40$ and the initial conditions for the Tobit Kalman filter and the particle filter are $\mathbf{x_0} = [\mathbf{-100}; \mathbf{0}; \mathbf{-100}; \mathbf{0}]$, with the true state being $\mathbf{x_k} = [\mathbf{50}; \mathbf{0}; \mathbf{50}; \mathbf{0}]$. The particle filter is biased in this example, as it converges to a value that is not $\mathbf{x_k}$, whereas the Tobit Kalman filter converges to the true value.

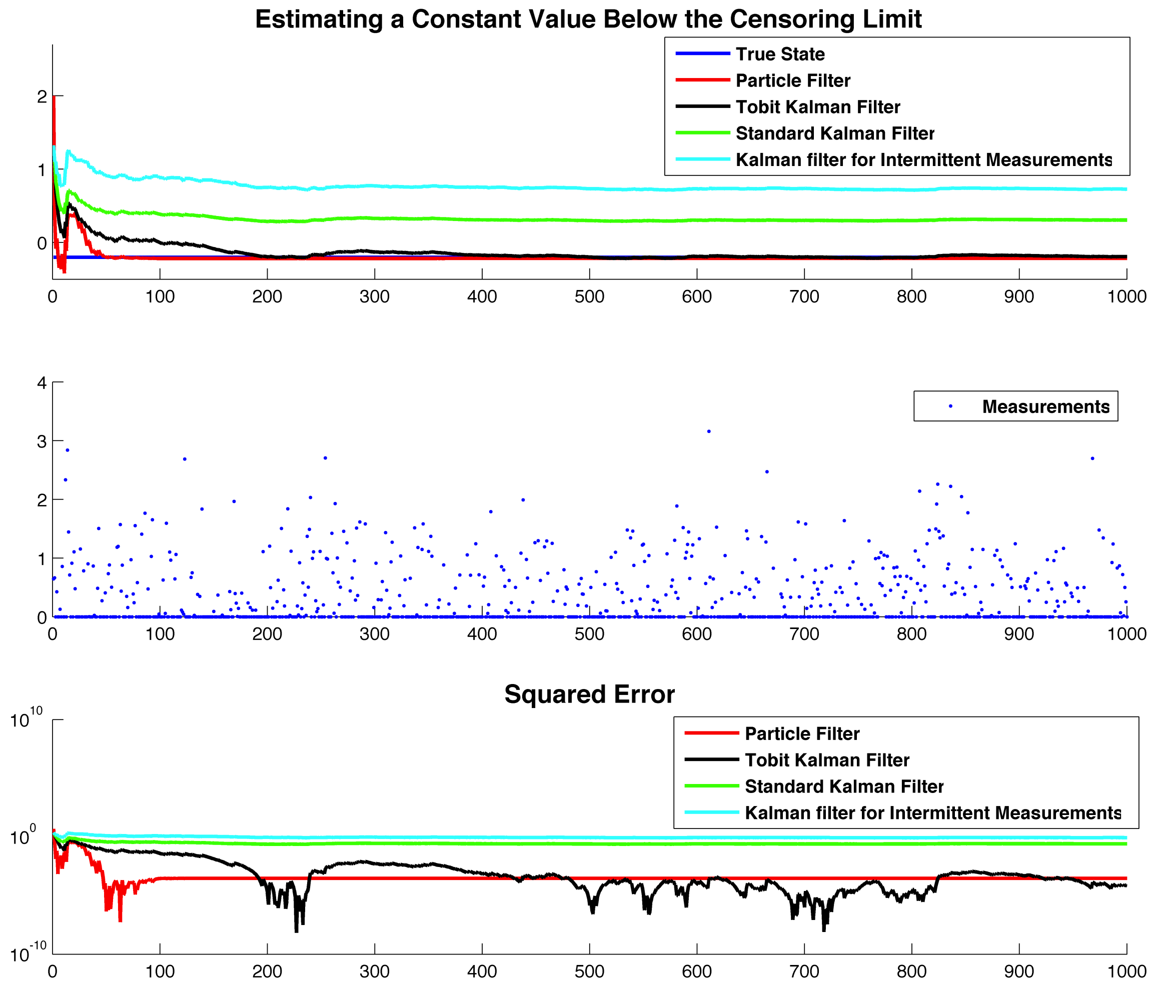

Figure 6.4: Experiment 2 with too few particles: Constant value below the censoring limit, the particle filter is run with 10 particles.

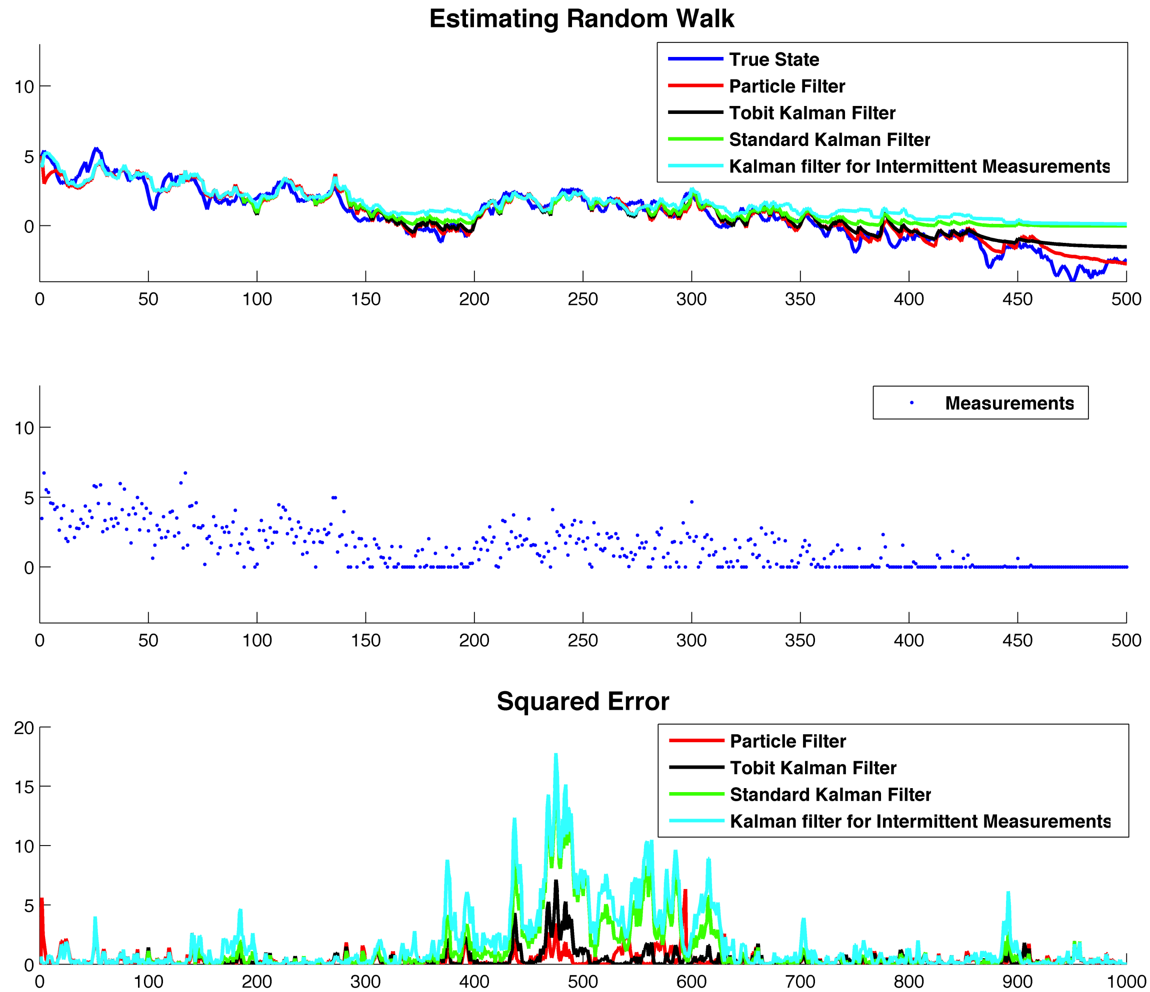

Figure 6.5: Comparison of particle filter, Tobit Kalman filter, SKF and KFIM in estimating a random walk signal near the censoring limit

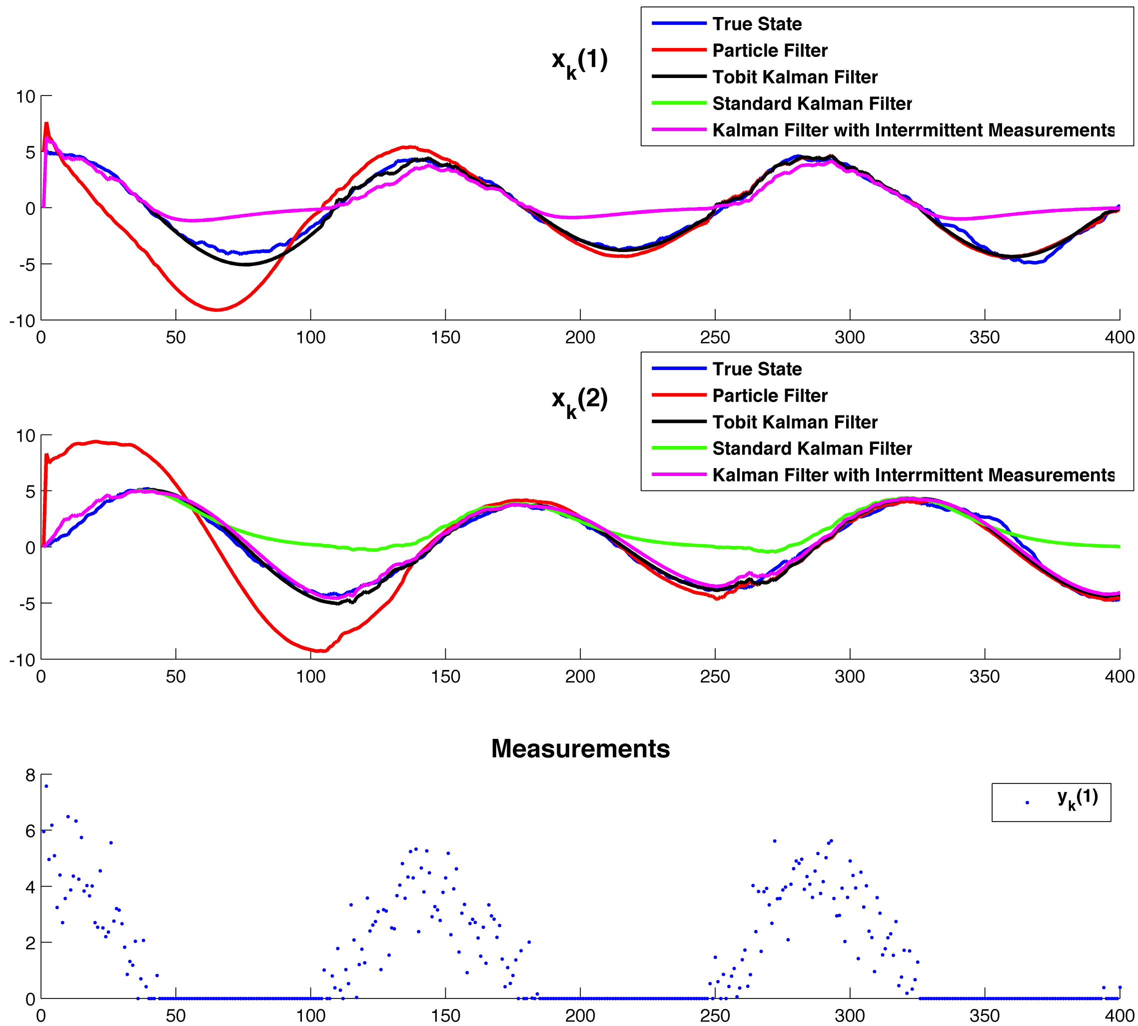

Figure 6.6: Comparison of particle filter, Tobit Kalman filter, SKF and KFIM in estimating a sinusoidal signal with small processes noise around the censoring limit

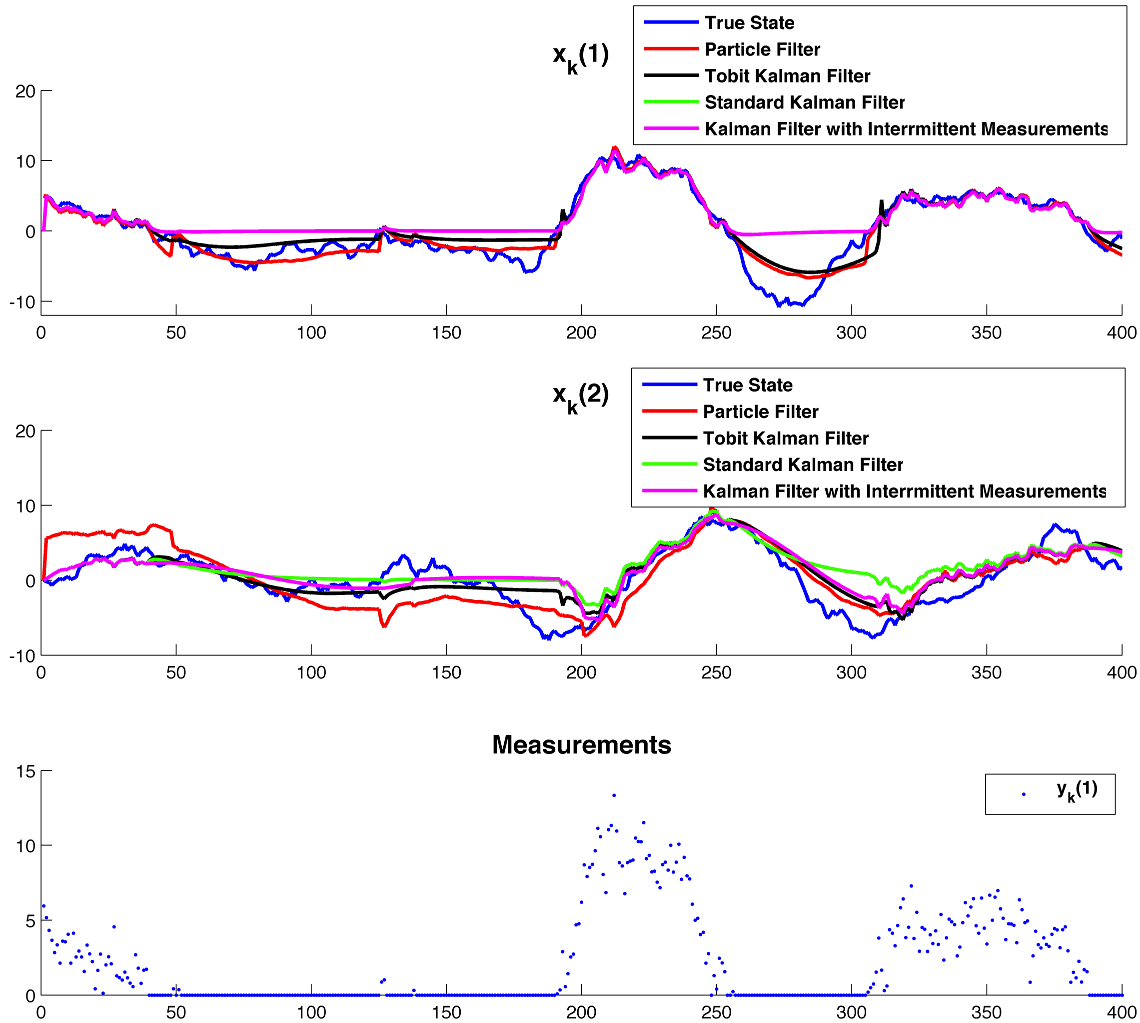

Figure 6.7: Comparison of particle filter, Tobit Kalman filter, SKF and KFIM in estimating a sinusoidal signal with large processes noise around the censoring limit

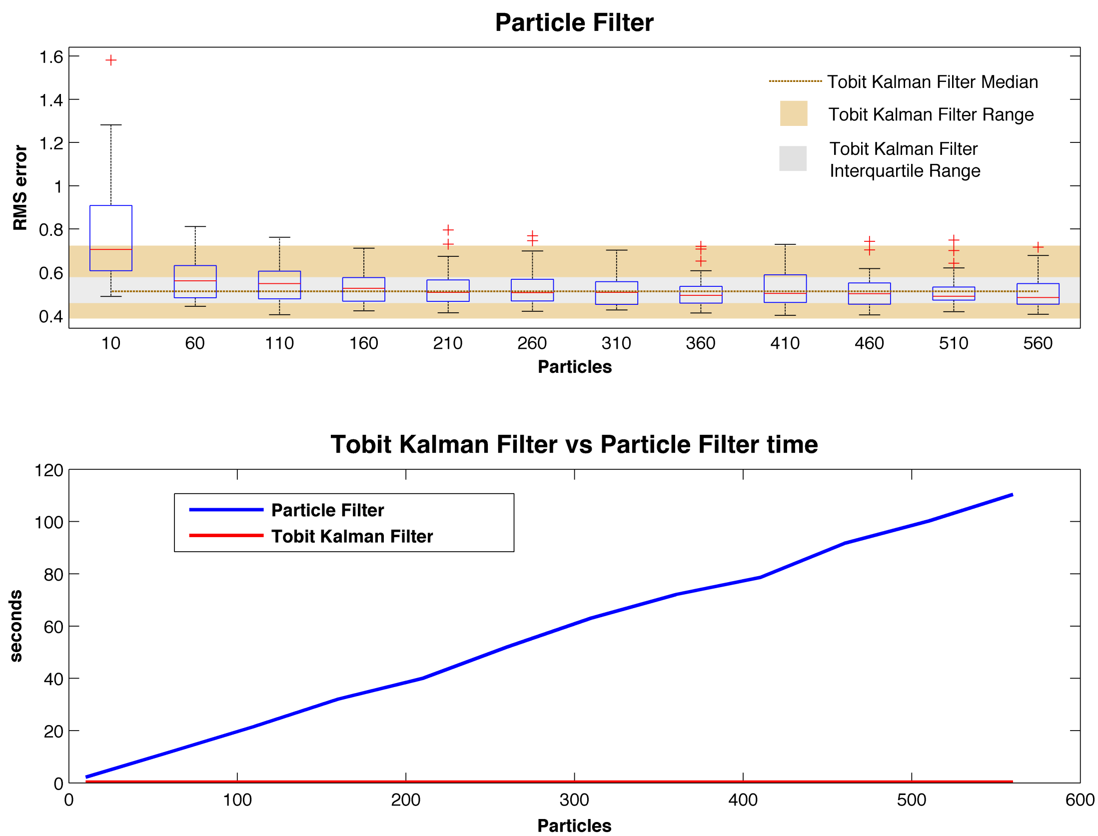

Figure 6.8: RMS error and time comparison of the Tobit Kalman filter and the particle filter. Statistics shown with 50 sets of simulated data.

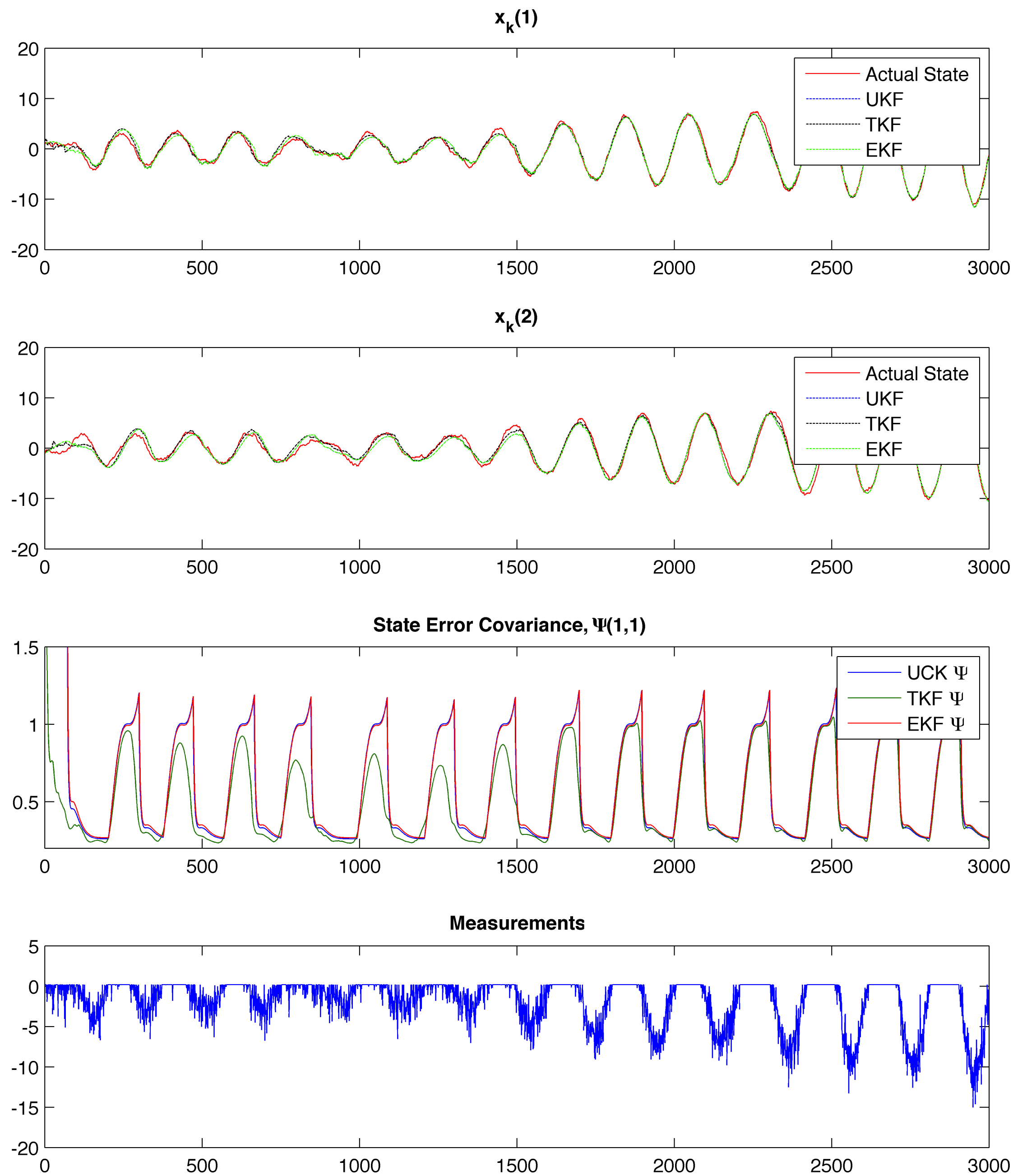

Figure 6.9: Sinusoidal example comparing the EKF, UKF and the Tobit Kalman filter. State estimates plotted with the state error covariance and measurements

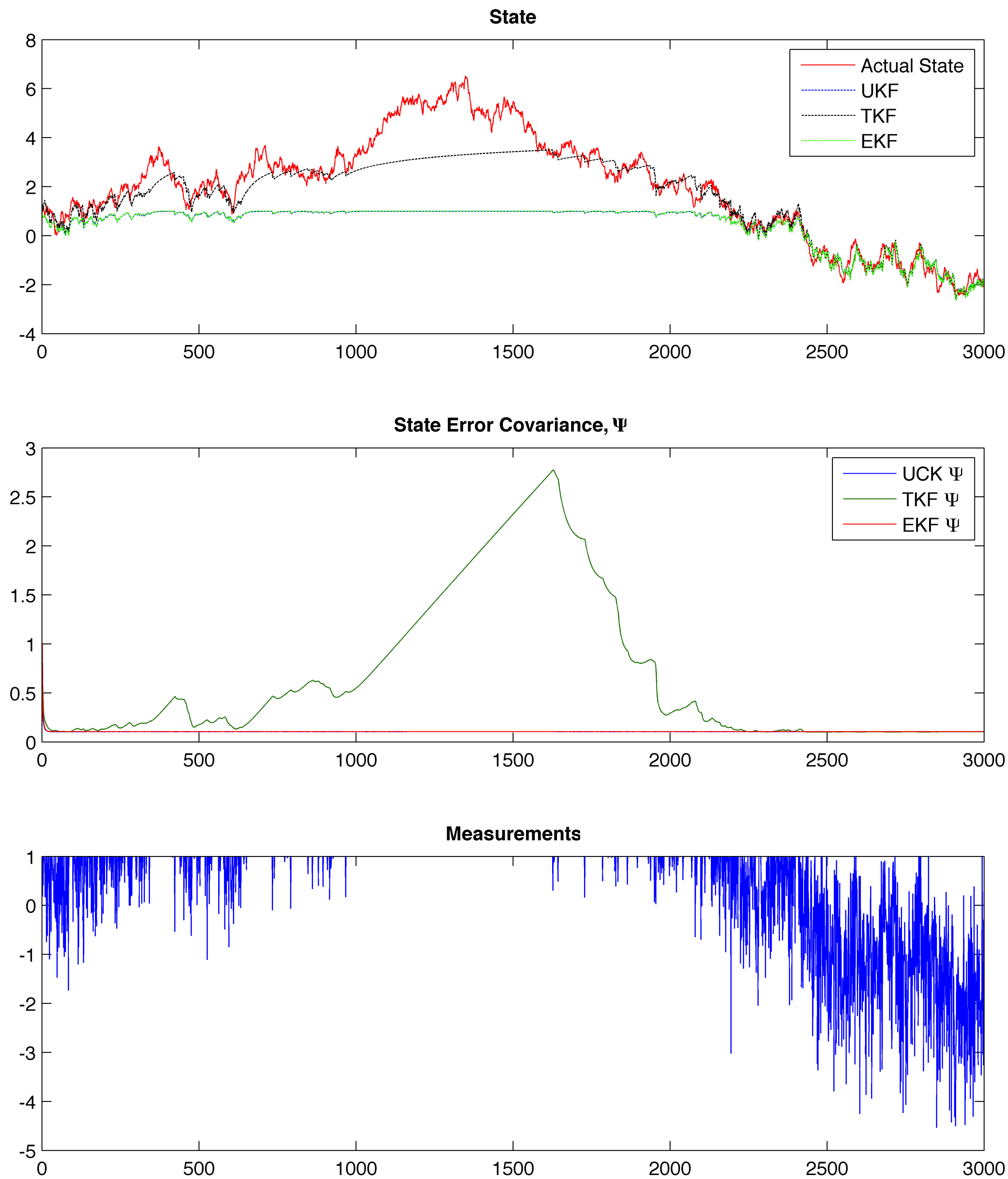

Figure 6.10: Random walk comparing the EKF, UKF and the Tobit Kalman filter. State estimates plotted with the state error covariance and measurements

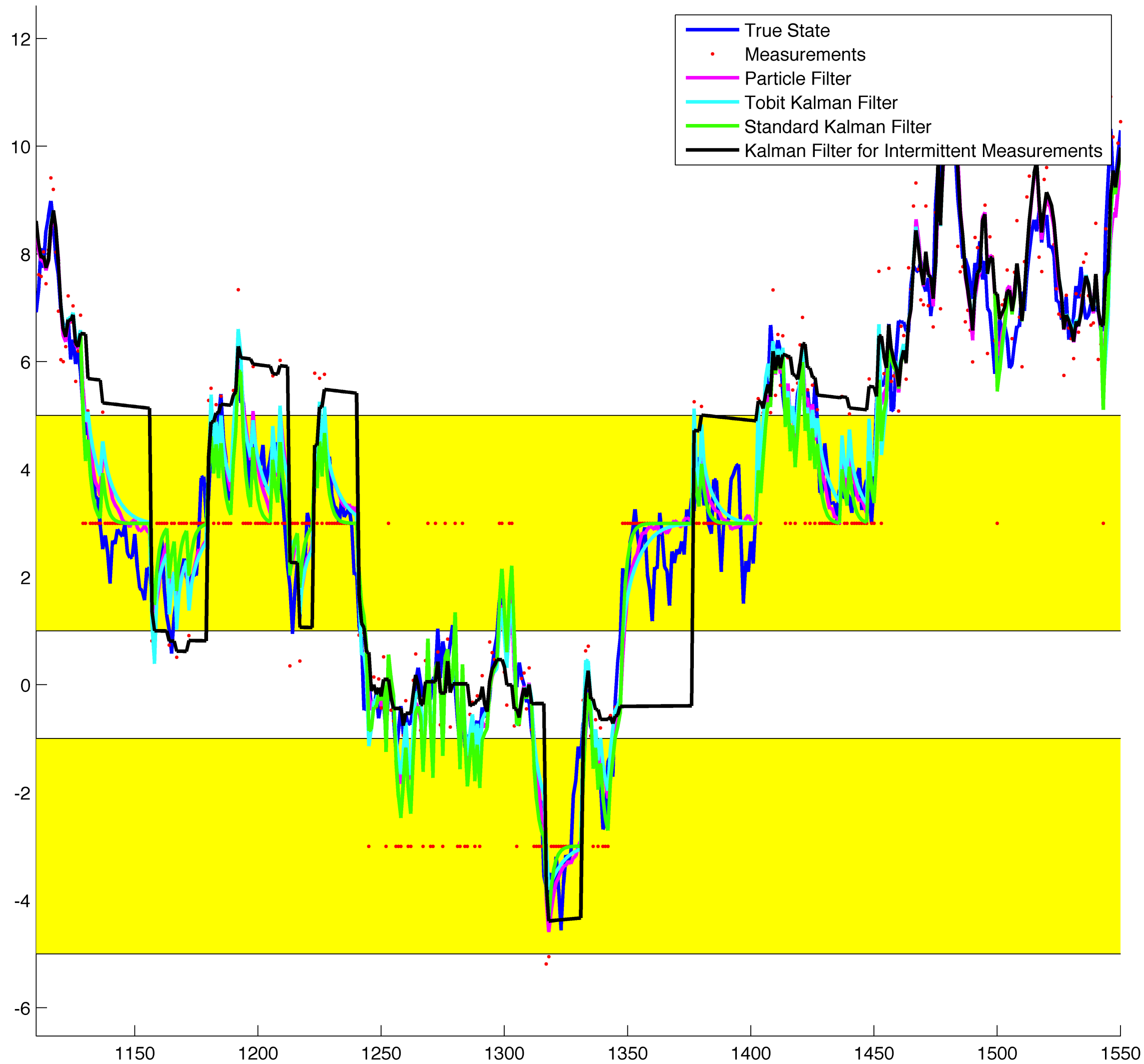

Figure 6.11: Random walk example comparing the particle filter, Tobit Kalman filter, SKF and KFIM with 1D occlusions

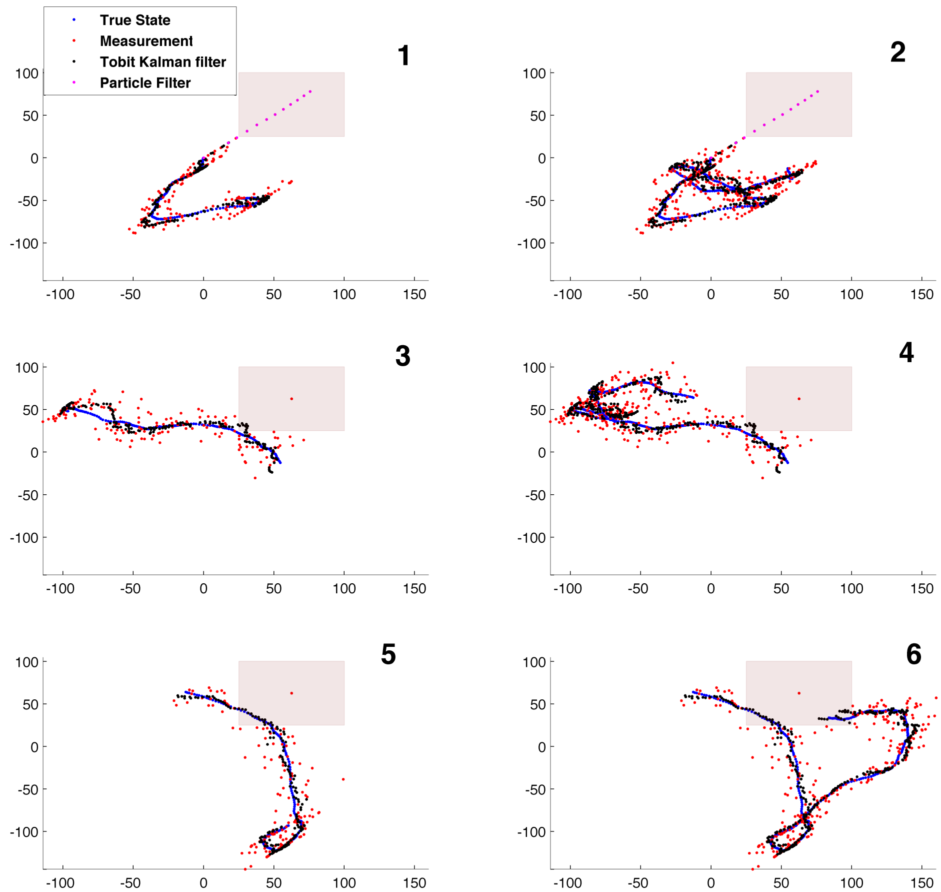

Figure 6.12: 2D spatial random walk comparing the particle filter and Tobit Kalman filter with one occlusion. Particle filter fails.

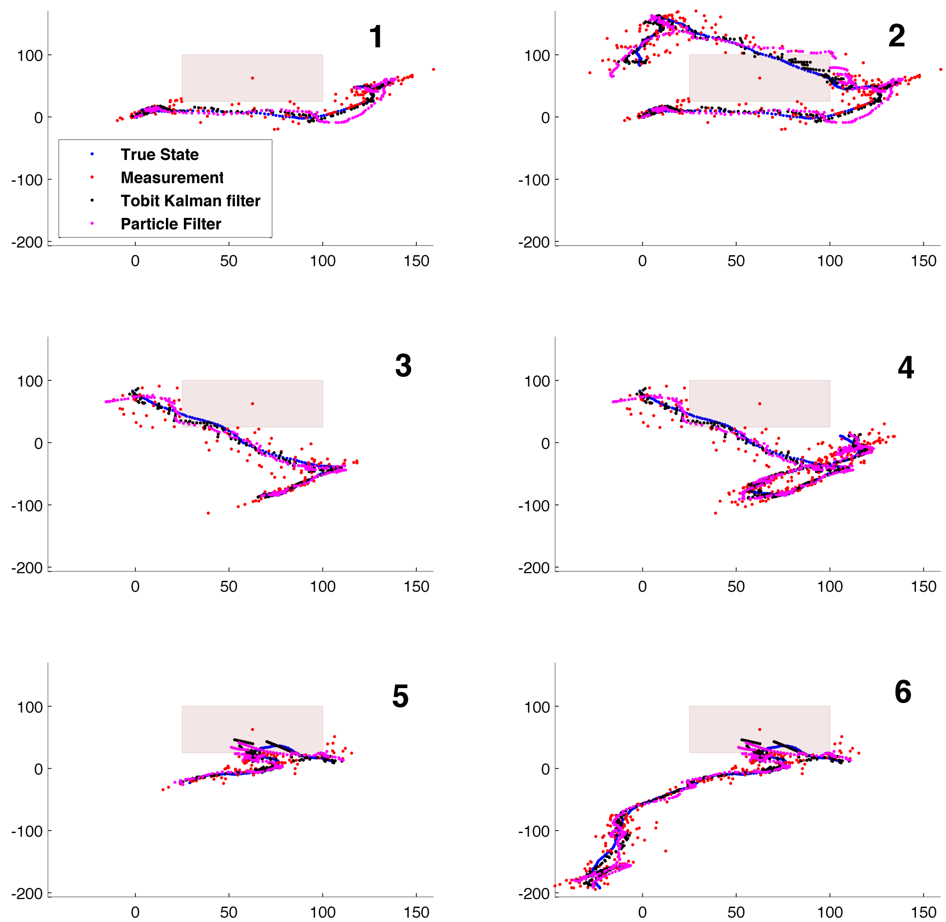

Figure 6.13: 2D spatial random walk comparing the particle filter and Tobit Kalman filter with one occlusion. Particle filter does not fail but suffers latency.

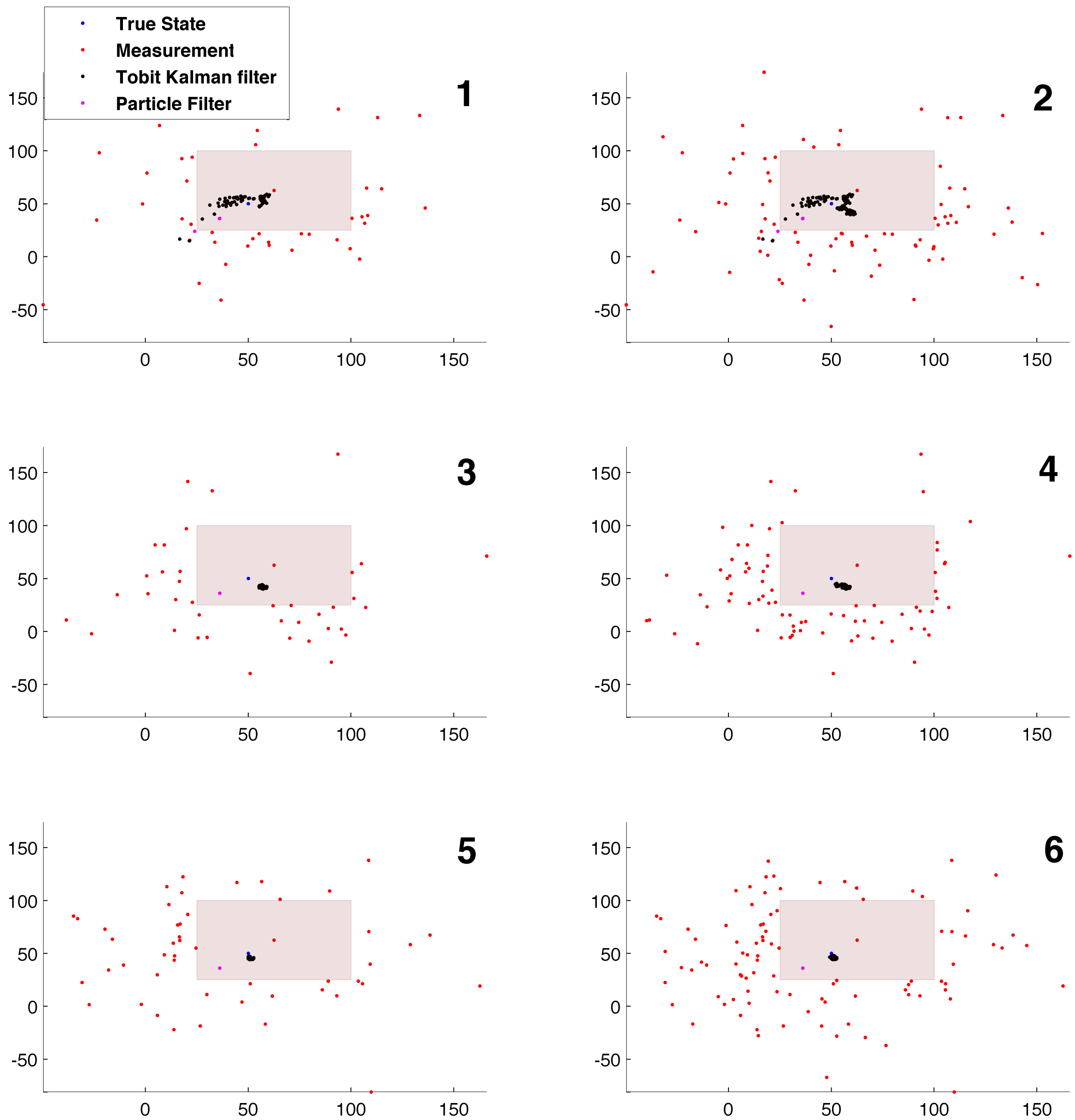

Figure 6.14: 2D spatial stationary system comparing the particle filter and Tobit Kalman filter with one occlusion.

Chapter 7

CONCLUSIONS AND FUTURE WORK

In this dissertation a novel adaptation of the Kalman filter for Tobit type 1 censored measurements is presented, which we call the Tobit Kalman filter. The resulting formulation provides an unbiased estimate of the state even when a high proportion of the measurements are censored. A linear, recursive estimator for censored data was derived using the assumption that the probability of the measurement being censored or not censored is predictable. The result of this assumption is that the *a priori* estimate of the state can be used to calculate the probability of being censored, and as a result, the Tobit Kalman filter innovation and gain.

The behavior of the Tobit Kalman filter is compared to five other methods: one in which the censored values are used as true measurements, another which treats the censored data points as missing, the extended Kalman filter, the unscented Kalman filter and the particle filter. The Tobit Kalman filter consistently outperforms four computationally equivalent filters in situations with moderate censoring. The particle filter is a preferred estimator for nonlinear models because it has an the advantage that there are no restrictions on the system model or the distribution. This advantage comes with a price of higher computational cost. The Tobit Kalman filter often performs better than the particle filter with much less computational requirements. The particle filter will perform well with a large number of particles and when there is disturbance in the model. However, the particle filter suffers from sample degeneracy, causing weights to collapse to one value.

Censoring is heavily studied in biology in the form of survival models and in economics with the Tobit model framework. In addition, censored data arises naturally in a number of engineering applications. Applications for the Tobit Kalman filter formulation include biochemical measurements with limit-of-detection saturation, inexpensive sensors with saturation censoring, visual tracking with camera frame censoring, and line-of-sight tracking with occlusion. In this dissertation, the Tobit Kalman filter is derived for left and right censoring, saturation and occlusions type censoring in 1D and 2D.

7.1 Future Work

The Tobit Kalman filter has the possibility of being implemented in many different applications including biology, control systems, computer vision based tracking, spatial localization etc. Also, building upon the Tobit type 1 model, one could design a Tobit Kalman filter for Tobit type 2,3,4 or 5 censoring. As stated in Chapter 1, there are several engineering applications for different types of Tobit censoring. For example, any system with state dependent censoring limits could be designed as a Tobit type 2 system. The 2D Tobit Kalman filter for spatial tracking has the heavy dependence in measurement noise due to the cross dependence in the x and y directions. This dependence is cause to pursue a Tobit Type 2 model, which allows for latent variable censoring dependence.

In the following section, the very important application of categorical data is briefly discussed. The categorical data problem occurs in biology and social sciences in the form of surveys or polls, and in engineering in the form of quantization.

7.1.1 Categorical Data

In many biological and social applications, categorical data is used to make inferences on public health [78]. The Tobit Kalman filter can be applied to categorical data if all the data is assumed to be censored and each censoring region is known. For N bin or categorical responses,The categorical model is,

$$
\mathbf{x_k} = \mathbf{A}\mathbf{x_{k-1}} + \mathbf{v_k}
$$

$$
\mathbf{y_k^*} = \mathbf{C}\mathbf{x_k} + \mathbf{w_k}
$$

$$
\mathbf{y_k} = \begin{cases} M(\infty, T_{high,1}), & \mathbf{Cx_k} + \mathbf{w_k} > T_{high,1} \\ M(T_{high,1}, T_{low,1}), & T_{low,1} < \mathbf{Cx_k} + \mathbf{w_k} < T_{high,1} \\ M(T_{low,1}, T_{high,2}), & T_{high,2} < \mathbf{Cx_k} + \mathbf{w_k} < T_{low,1} \\ M(T_{high,2}, T_{low,2}), & T_{low,2} < \mathbf{Cx_k} + \mathbf{w_k} < T_{high,2} \\ \vdots \\ M(T_{low,N}, -\infty), & \mathbf{Cx_k} + \mathbf{w_k} < T_{low,N} \end{cases}
$$

Where $M(\alpha, \beta)$ represents the censored measurement when the latent measurement is between the values α and β. The probability distribution of the measurement is,

$$
f_c(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) = \begin{cases} \Delta_{k,1} \int_{T_{low,1}}^{\infty} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) d\mathbf{y_k}, & \mathbf{y_k^*} > T_{high,1} \\ \Delta_{k,1} \int_{T_{low,1}}^{T_{high,1}} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} & T_{low,1} < \mathbf{y_k^*} < T_{high,1} \\ \Delta_{k,1} \int_{T_{high,2}}^{T_{low,1}} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) d\mathbf{y_k}, & T_{high,2} < \mathbf{y_k^*} < T_{low,1} \\ \Delta_{k,2} \int_{T_{low,2}}^{T_{high,2}} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} & T_{low,2} < \mathbf{y_k^*} < T_{high,2} \\ \vdots \\ \Delta_{k,1} \int_{-\infty}^{T_{high,1}} f(\mathbf{y_k}|\mathbf{R}, \mathbf{x_{k|k-1}}) d\mathbf{y_k} & \mathbf{y_k^*} < T_{low,N} \end{cases}
$$

With $\Delta_{k,1} = \delta(\mathbf{y_k} - M(T_{high,1}, T_{low,1}))$, the Kronecker delta function.

For a binary measurement application, such as a yes/no questionnaire where the latent variable is continuous the Tobit Kalman estimator has the possibility to give better estimates to answers of "how are you feeling today?" or "What color do you see, red or blue?" when the answers are binary and a system model is given. The measurement model contains two bins of categorical responses represented as,

$$
\mathbf{y_k} = \begin{cases} M(\infty, T_{high,1}), & \mathbf{Cx_k} + \mathbf{w_k} > T \\ M(T_{low,N}, -\infty), & \mathbf{Cx_k} + \mathbf{w_k} < T \end{cases}
$$

With distribution,

$$f_c(\mathbf{y_k}|\mathbf{R},\mathbf{x_{k|k-1}}) = \begin{cases} \Delta_{k,1}\int_T^\infty f(\mathbf{y_k}|\mathbf{R},\mathbf{x_{k|k-1}})d\mathbf{y_k}, & \mathbf{y_k^*} > T \\ \Delta_{k,2}\int_{-\infty}^T f(\mathbf{y_k}|\mathbf{R},\mathbf{x_{k|k-1}})d\mathbf{y_k} & \mathbf{y_k^*} < T \end{cases}$$

The expected value of the measurement given the state estimate is,

$$E(y_k) = P(Cx_k > T)E(y_k|Cx_k > T) + P(Cx_k < T)E(y_k|Cx_k < T)$$

Where $E(y_k|Cx_k > T)M(T_{low,N}, -\infty)$ and $E(y_k|Cx_k > T) = M(\infty, T_{high,1})$ are the censored measurements or binary values. The numeric value representing a yes/no response does not have a censored data value as with the one sided censoring, so finding the Tobit Kalman gain, and innovation will be difficult. The complete formulation for the categorical data is left for future work.

BIBLIOGRAPHY

[1] J.F McDonald and R.A Moffitt. The uses of tobit analysis. *The review of economics and statistics*, 62(2):318–321, 1980.

[2] J Tobin. Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, pages 24–36, 1958.

[3] J Adams. Personal wealth transfers. *The Quarterly Journal of Economics*, Jan 1980.

[4] S Stephenson and J McDonald. Disaggregation of income maintenance impacts on family earnings. *Review of Economics and Statistics*, Jan 1979.

[5] A Adesina and M Zinnah. Technology characteristics, farmers' perceptions and adoption decisions: A tobit model application in sierra leone. *Agricultural economics*, Jan 1993.

[6] T Amemiya. Advanced econometrics. *Harvard university press*, Jan 1985.

[7] R Singh and K Mukhopadhyay. Survival analysis in clinical trials: Basics and must know areas. *Perspectives in clinical research*, Jan 2011.

[8] K Leung, R Elashoff, and A Afifi. Censoring issues in survival analysis. *Annual Review of Public Health*, Jan 1997.

[9] H Pham. Springer handbook of engineering statistics. *Springer*, Jan 2006.

[10] S Prinja, N Gupta, and R Verma. Censoring in clinical trials: Review of survival analysis techniques. *... of Indian Association of Preventive & ...*, Jan 2010.

[11] V De Gruttola and S Lagakos. Analysis of doubly-censored survival data, with application to aids. *Biometrics*, Jan 1989.

[12] D Wang and A Bakhai. Clinical trials: A practical guide to design, analysis, and reporting. *Remedica*, Jan 2006.

[13] S Richards. A handbook of parametric survival models for actuarial use. *Scandinavian Actuarial Journal*, Jan 2012.

[14] T Amemiya. Regression analysis when the dependent variable is truncated normal. *Econometrica: Journal of the Econometric Society*, pages 997–1016, 1973.

[15] F Giri and E Bai. Block-oriented nonlinear system identification. *Springer*, Jan 2010.

[16] T Wigren. Convergence analysis of recursive identification algorithms based on the nonlinear wiener model. *Automatic Control*, Jan 1994.

[17] JB Hampshire and J.W Strohbehn. Tobit maximum-likelihood estimation for stochastic time series affected by receiver saturation. *Information Theory, IEEE Transactions on*, 38(2):457–469, 1992.

[18] T Amemiya. The estimation of a simultaneous-equation tobit model. *International Economic Review*, pages 169–181, 1979.

[19] R.T Carson and Y Sun. The tobit model with a nonzero threshold. *The Econometrics Journal*, 10(3):488–502, 2007.

[20] R.J Smith and R.W Blundell. An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica: Journal of the Econometric Society*, pages 679–685, 1986.

[21] W Schnedler. Likelihood estimation for censored random vectors. *Econometric Reviews*, 24(2):195–217, 2005.

[22] Bethany Allik, Cory Miller, Mike Piovoso, and Ryan Zurakowski. Estimation of saturated data using the tobit kalman filter. *American Control Conference*, 2014.

[23] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82:35–45, 1960.

[24] B Ristic, S Arulampalam, and N Gordon. Beyond the kalman filter: Particle filters for tracking applications. *Artech house*, Jan 2004.

[25] B Sinopoli, L Schenato, M Franceschetti, K Poolla, M.I Jordan, and S.S Sastry. Kalman filtering with intermittent observations. *Automatic Control, IEEE Transactions on*, 49(9):1453–1464, 2004.

[26] S Kar, B Sinopoli, and J.M.F Moura. Kalman filtering with intermittent observations: Weak convergence to a stationary distribution. *Automatic Control, IEEE Transactions on*, 57(2):405–420, 2012.

[27] S.C Smith and P Seiler. Estimation with lossy measurements: jump estimators for jump systems. *Automatic Control, IEEE Transactions on*, 48(12):2163–2171, 2003.

[28] B Ibarz-Gabardos and P J Zufiria. A kalman filter with censored data. *Intelligent Signal Processing, 2005 IEEE International Workshop on*, pages 74–79, 2005.

[29] Yujiao Zheng, Ruixin Niu, and P K Statistical Signal Processing Workshop SSP 2012 IEEE Varshney. Sequential bayesian estimation with censored data. *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, 2012.

[30] Constantino Rago, Peter Willett, and Yaakov Bar-Shalom. Censoring sensors: A low-communication-rate scheme for distributed detection. *Aerospace and Electronic Systems, IEEE Transactions on*, 32(2):554–568, 1996.

[31] Davide Spinello and Daniel J Stilwell. Nonlinear estimation with state-dependent gaussian observation noise. *IEEE Transactions on Automatic Control*, 55:1358–1366, 2010.

[32] Christophe Andrieu and Arnaud Doucet. Particle filtering for partially observed gaussian state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):827–836, 2002.

[33] J Heckman and E Leamer. Handbook of econometrics. *Elsevier*, Jan 2001.

[34] James o Berger. Statistical decision theory and bayesian analysis. *Springer*, 1985.

[35] G Bishop and G Welch. An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8:27599–23175, 2001.

[36] S Särkkä. Bayesian filtering and smoothing. *Cambridge University Press*, 3, Jan 2013.

[37] Z Chen. Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, Jan 2003.

[38] Dan Simon. Optimal state estimation: Kalman, h infinity, and nonlinear approaches. *John Wiley & Sons*, 2006.

[39] Louis L Scharf. Statistical signal processing detection, estimation, and time series analysis. *REading, MA: Addison-Wesley*, 98, 1991.

[40] D Simon. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory & Applications*, 4(8):1303–1318, Aug 2010.

[41] Bradford Davis, Michael Hamilton, and David Hepner. Shock experiment results of the dfuze 8-channel inertial sensor suite that contains commercial magnetometers and accelerometers. *ARL Technical Report*, 2002.

[42] Mark Ilg. Guidance, navigation, and control for munitions. *Doctoral Dissertation, Drexel University*, 2008.

[43] Andrew Thompson. Calibration of inertial sensors. *ARL Technical Report*, 2000.

[44] R Kandepu, B Foss, and L Imsland. Applying the unscented kalman filter for nonlinear state estimation. *Journal of Process Control*, Jan 2008.

[45] S Haykin, S Haykin, and S Haykin. Kalman filtering and neural networks. *Wiley Online Library*, Jan 2001.

[46] M St-Pierre and D Gingras. Comparison between the unscented kalman filter and the extended kalman filter for the position estimation module of an integrated navigation information system. *Intelligent Vehicles Symposium*, Jan 2004.

[47] B Allik, M Ilg, and R Zurakowski. Ballistic roll estimation using ekf frequency tracking and adaptive noise cancellation. *Aerospace and Electronic Systems ...*, Jan 2013.

[48] D Tenne and T Singh. The higher order unscented filter. *... Conference*, Jan 2003.

[49] E Wan and R Van Der Merwe. The unscented kalman filter for nonlinear estimation. *... 2000. AS-SPCC. The IEEE 2000*, Jan 2000.

[50] A Romanenko and J Castro. The unscented filter as an alternative to the ekf for nonlinear state estimation: a simulation case study. *Computers & chemical engineering*, Jan 2004.

[51] S Julier and J Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, Jan 2004.

[52] J Tugnait. Detection and estimation for abruptly changing systems. *Automatica*, Jan 1982.

[53] Cory Miller Mike Piovoso Ryan Zurakowski Bethany Allik. The tobit kalman filter: An estimator for censored data. *IEEE Transactions on Control Systems Technology*, page (submitted), 2013.

[54] C Snyder, T Bengtsson, and P Bickel.... Obstacles to high-dimensional particle filtering. *Monthly Weather ...*, Jan 2008.

[55] R Douc and O Cappé. Comparison of resampling schemes for particle filtering. *Image and Signal Processing and Analysis*, Jan 2005.

[56] A Doucet, S Godsill, and C Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, Jan 2000.

[57] A Doucet. Sequential monte carlo methods. *Wiley Online Library*, Jan 2001.

[58] G Calzolari and G Fiorentini. Estimating variances and covariances in a censored regression model. *Statistica*, 53:323–339, 1993.

[59] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm Computing Surveys (CSUR)*, 38(4):13, 2006.

[60] S Chen. Kalman filter for robot vision: a survey. *Industrial Electronics*, Jan 2012.

[61] Z Kalal, K Mikolajczyk, and J Matas. Tracking-learning-detection. *Pattern Analysis and Machine . . .*, Jan 2012.

[62] Miller C, B Allik, M Ilg, and R Zurakowski. Kalman filter-based tracking of multiple similar objects from a moving camera platform. pages 5679–5684, 2012.

[63] S Ali and M Shah. Cocoa: tracking in aerial imagery. *Defense and Security Symposium*, Jan 2006.

[64] V Dobrokhodov, I Kaminer, and K Jones. . . . Vision-based tracking and motion estimation for moving targets using small uavs. *American Control . . .*, Jan 2006.

[65] T Diwan, F Lone, and V Sahu. Robust tracking and object classification towards automated video surveillance recognition. *Digital Image Processing*, Jan 2013.

[66] M Isard and A Blake. Contour tracking by stochastic propagation of conditional density. *Computer Vision—ECCV'96*, Jan 1996.

[67] A Yilmaz, X Li, and M Shah. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *Pattern Analysis and Machine . . .*, Jan 2004.

[68] V Pavlovic, R Sharma, and T Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *Pattern Analysis and . . .*, Jan 1997.

[69] D Beymer, P McLauchlan, and B Coifman. . . . A real-time computer vision system for measuring traffic parameters. *Computer Vision and . . .*, Jan 1997.

[70] C Schlegel, J Illmann, H Jaberg, M Schuster, and R Wörz. Vision based person tracking with a mobile robot. *BMVC*, Jan 1998.

[71] N Papanikolopoulos and P Khosla. . . . Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision. *. . . and Automation*, Jan 1993.

[72] Gregory Hager and Peter Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(10):1025–1039, 1998.

[73] Pierre Gabriel, Jacques Verly, Justus Piater, and André Genon. The state of the art in multiple object tracking under occlusion in video sequences. *Advanced Concepts for Intelligent Vision Systems*, pages 166–173, 2003.

[74] Cuevas, Erik V, Zaldivar, Daniel, Rojas, and Raul. Kalman filter for vision tracking. *Freie Univ., Fachbereich Mathematik und Informatik*, 2005.

[75] M Sanjeev Arulampalam and S Maskell.... A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing*, Jan 2002.

[76] Bethany Allik, Cory Miller, Mike Piovoso, and Ryan Zurakowski. Nonlinear estimators for censored data: A comparison of the ekf, the ukf and the tobit kalman filter. *Conference on Decision and Control*, page (submitted), 2015.

[77] Cory Miller, Bethany Allik, Mike Piovoso, and Ryan Zurakowski. Estimation of a mobile vehicle range and position using the tobit kalman filter. *Conference on Decision and Control*, page (submitted), 2014.

[78] J Preisser and G Koch. Categorical data analysis in public health. *Annual Review of Public Health*, Jan 1997.

Appendix A

USEFUL INTEGRALS

In this section we will list some useful integrals for Tobit Model statistics. The relationship between the probability distribution and the cumulative distribution function is,

$$\int_A^B \phi(x)\,\mathrm{d}x = \Phi(B) - \Phi(A)$$

The following two integrals are used to calculate the mean and variance,

$$\int x\phi(\frac{x-\mu_x}{\sigma})\,\mathrm{d}x = -\sigma^2\phi(\frac{x-\mu_x}{\sigma}) + \sigma\mu_x\Phi(\frac{x-\mu_x}{\sigma})$$

$$\int x^2\phi(\frac{x-\mu_x}{\sigma})\,\mathrm{d}x = \sigma^2(\mu_x^2+1)\Phi(\frac{x-\mu_x}{\sigma}) - \sigma^2(\mu_x+x)\phi(\frac{x-\mu_x}{\sigma})$$

For the two dimensional case, since the two dimemsions are independent we have the useful property $\phi(x,y) = \phi(x)\phi(y)$, so that,

$$\int_A^B\int_C^D \phi(x,y)\,\mathrm{d}x\,\mathrm{d}y = \Phi(D) - \Phi(C)\int_A^B \phi(x)\,\mathrm{d}x = (\Phi(D) - \Phi(C))(\Phi(B) - \Phi(A))$$

A, B, C and D do not depend on x, y.

Appendix B

DERIVATION OF $\mathbf{R_{\tilde{Y}\tilde{Y}_K}}$

The derivation of $\mathbf{R_{\tilde{y}\tilde{y}_k}}$ is straight forward given assumption 1. Recall,

$$\mathbf{R_{\tilde{y}\tilde{y}_k}} = \mathbf{E}(\mathbf{\tilde{y}_k}\mathbf{\tilde{y}_k^T}) \tag{B.1}$$

and that,

$$\mathbf{\tilde{y}_k} = \mathbf{p_k}(\mathbf{C}\mathbf{x_k} + \mathbf{v_k}) + (\mathbf{I_{m\times m}} - \mathbf{p_k})\mathbf{\Gamma} - \mathbf{E}(\mathbf{y_k}) \tag{B.2}$$

The value of $\mathbf{E}(\mathbf{y_k})$ is,

$$\mathbf{E}(\mathbf{y_k}) = \mathbf{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}} - \mathbf{\Gamma}, \mathbf{\Sigma})[\mathbf{C}\mathbf{x_{k|k-1}} + \mathbf{\Sigma}\mathbf{\Lambda}(\mathbf{\Gamma} - \mathbf{C}\mathbf{x_{k|k-1}}, \mathbf{\Sigma})] + \mathbf{\Phi}(\mathbf{\Gamma} - \mathbf{C}\mathbf{x_{k|k-1}}, \mathbf{\Sigma})\mathbf{\Gamma} \tag{B.3}$$

Where,

$$\mathbf{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}} - \mathbf{\Gamma}, \mathbf{\Sigma}) = \mathbf{Diag}\begin{pmatrix} \Phi(\frac{Cx_{k|k-1}(1)-\gamma(1)}{\sigma(1)}) \\ \Phi(\frac{Cx_{k|k-1}(2)-\gamma(2)}{\sigma(2)}) \\ \vdots \\ \Phi(\frac{Cx_{k|k-1}(m)-\gamma(m)}{\sigma(m)}) \end{pmatrix}. \tag{B.4}$$

and $\mathbf{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}} - \mathbf{\Gamma}, \mathbf{\Sigma}) = \mathbf{I_{m\times m}} - \mathbf{\Phi}(\mathbf{\Gamma} - \mathbf{C}\mathbf{x_{k|k-1}}, \mathbf{\Sigma})$.

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(m) \end{pmatrix}. \tag{B.5}$$

$\mathbf{\Sigma}$ is a diagonal matrix representing the measurement noise for each measurement $\mathbf{y_k}$, without censoring.

$$\mathbf{\Sigma} = \mathbf{Diag}\begin{pmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(m) \end{pmatrix}. \tag{B.6}$$

The matrix form of the inverse mills ratio,

$$\mathbf{\Lambda}(\mathbf{\Gamma} - \mathbf{Cx_k}, \mathbf{\Sigma}) = \begin{pmatrix} \lambda(\gamma(1) - Cx_{k|k-1}(1), \sigma(1)) \\ \lambda(\gamma(2) - Cx_{k|k-1}(2), \sigma(2)) \\ \vdots \\ \lambda(\gamma(m) - Cx_{k|k-1}(m), \sigma(m)) \end{pmatrix}. \tag{B.7}$$

Using the expanded version of variance, that $E((x - E(x))^2) = E(x^2) - E(x)^2$, $\mathbf{E}(\mathbf{p_k}(\mathbf{Cx_k} + \mathbf{v_k}) + (\mathbf{I_{m\times m}} - \mathbf{p_k})\mathbf{\Gamma}) = \mathbf{E}(\mathbf{y_k})$ we can reduce Equation B.1 to,

$$\begin{aligned} \mathbf{R_{\tilde{y}\tilde{y}_k}} &= \mathbf{E}((\mathbf{p_k}(\mathbf{Cx_k} + \mathbf{v_k}) + (\mathbf{I_{m\times m}} - \mathbf{p_k})\mathbf{\Gamma})(\mathbf{p_k}(\mathbf{Cx_k} + \mathbf{v_k}) + (\mathbf{I_{m\times m}} - \mathbf{p_k})\mathbf{\Gamma})^{\mathbf{T}}) \\ &\quad -\mathbf{E}(\mathbf{y_k})\mathbf{E}(\mathbf{y_k})^{\mathbf{T}} \end{aligned} \tag{B.8}$$

We will next derive the terms in Equation B.8, $startingwith\mathbf{E}(\mathbf{y_k})\mathbf{E}(\mathbf{y_k})^{\mathbf{T}}$,

$$\begin{aligned} &\mathbf{E}(\mathbf{y_k})\mathbf{E}(\mathbf{y_k})^{\mathbf{T}} = \\ &\mathbf{\Phi}(\mathbf{Cx_{k|k-1}} - \mathbf{\Gamma}, \mathbf{\Sigma})(\mathbf{Cx_{k|k-1}} + \mathbf{\Sigma\Lambda}(\mathbf{\Gamma} - \mathbf{Cx_{k|k-1}}, \mathbf{\Sigma})) \\ &(\mathbf{Cx_{k|k-1}} + \mathbf{\Sigma\Lambda}(\mathbf{\Gamma} - \mathbf{Cx_{k|k-1}}, \mathbf{\Sigma}))^{\mathbf{T}}\mathbf{\Phi}(\mathbf{Cx_{k|k-1}} - \mathbf{\Gamma}, \mathbf{\Sigma}) \\ &+\mathbf{\Phi}(\mathbf{Cx_{k|k-1}} - \mathbf{\Gamma}, \mathbf{\Sigma})(\mathbf{Cx_{k|k-1}} + \mathbf{\Sigma\Lambda}(\mathbf{\Gamma} - \mathbf{Cx_{k|k-1}}, \mathbf{\Sigma}))\mathbf{\Gamma}^{\mathbf{T}}\mathbf{\Phi}(\mathbf{\Gamma} - \mathbf{Cx_{k|k-1}}, \mathbf{\Sigma}) \\ &+\mathbf{\Phi}(\mathbf{\Gamma} - \mathbf{Cx_{k|k-1}}, \mathbf{\Sigma})\mathbf{\Gamma}(\mathbf{Cx_{k|k-1}} + \mathbf{\Sigma\Lambda}(\mathbf{\Gamma} - \mathbf{Cx_k}, \mathbf{\Sigma}))^{\mathbf{T}}\mathbf{\Phi}(\mathbf{Cx_{k|k-1}} - \mathbf{\Gamma}, \mathbf{\Sigma}) \\ &+\mathbf{\Phi}(\mathbf{\Gamma} - \mathbf{Cx_{k|k-1}}, \mathbf{\Sigma})\mathbf{\Gamma}\mathbf{\Gamma}^{\mathbf{T}}\mathbf{\Phi}(\mathbf{\Gamma} - \mathbf{Cx_{k|k-1}}, \mathbf{\Sigma}) \end{aligned} \tag{B.9}$$

and the first term of Equation B.8 is,

$$\begin{aligned}
&(\mathbf{p_k}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})+(\mathbf{I_{m\times m}}-\mathbf{p_k})\boldsymbol{\Gamma})(\mathbf{p_k}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})+(\mathbf{I_{m\times m}}-\mathbf{p_k})\boldsymbol{\Gamma})^{\mathbf{T}} = \\
&(\mathbf{p_k}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})^{\mathbf{T}}\mathbf{p_k}+\mathbf{p_k}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})\boldsymbol{\Gamma}^{\mathbf{T}}(\mathbf{I_{m\times m}}-\mathbf{p_k}) \\
&+(\mathbf{I_{m\times m}}-\mathbf{p_k})\boldsymbol{\Gamma}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})^{\mathbf{T}}\mathbf{p_k} \\
&+(\mathbf{I_{m\times m}}-\mathbf{p_k})\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathbf{T}}(\mathbf{I_{m\times m}}-\mathbf{p_k})
\end{aligned} \tag{B.10}$$

using assumption 1, and,

$$\mathbf{E}(\mathbf{p_k}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})) = \boldsymbol{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}}-\boldsymbol{\Gamma},\boldsymbol{\Sigma})(\mathbf{C}\mathbf{x_{k|k-1}}+\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}}\boldsymbol{\Sigma})) \tag{B.11}$$

$$\begin{aligned}
&\mathbf{E}(\mathbf{p_k}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})\boldsymbol{\Gamma}^{\mathbf{T}}(\mathbf{I_{m\times m}}-\mathbf{p_k})) = \\
&\boldsymbol{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}}-\boldsymbol{\Gamma},\boldsymbol{\Sigma})(\mathbf{C}\mathbf{x_{k|k-1}}+\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma}))\boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Phi}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_k},\boldsymbol{\Sigma})
\end{aligned} \tag{B.12}$$

$$\begin{aligned}
&\mathbf{E}((\mathbf{I_{m\times m}}-\mathbf{p_k})\boldsymbol{\Gamma}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})^{\mathbf{T}}\mathbf{p_k}) = \\
&\boldsymbol{\Phi}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma})\boldsymbol{\Gamma}(\mathbf{C}\mathbf{x_{k|k-1}}+\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma}))^{\mathbf{T}}\boldsymbol{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}}-\boldsymbol{\Gamma},\boldsymbol{\Sigma})
\end{aligned} \tag{B.13}$$

$$\begin{aligned}
&\mathbf{E}((\mathbf{I_{m\times m}}-\mathbf{p_k})\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathbf{T}}(\mathbf{I_{m\times m}}-\mathbf{p_k})) = \\
&\boldsymbol{\Phi}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma})\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Phi}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma})
\end{aligned} \tag{B.14}$$

Subtracting Equations B.9 and B.10, and using Equations B.12,B.13,B.14 we are left with,

$$\begin{aligned}
\mathbf{R_{\tilde{y}\tilde{y}_k}} &= \mathbf{E}(\mathbf{p_k}(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})(\mathbf{C}\mathbf{x_k}+\mathbf{v_k})^{\mathbf{T}}\mathbf{p_k}) \\
&-\boldsymbol{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}}-\boldsymbol{\Gamma},\boldsymbol{\Sigma})(\mathbf{C}\mathbf{x_{k|k-1}}+\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma})) \\
&(\mathbf{C}\mathbf{x_{k|k-1}}+\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma}))^{\mathbf{T}}\boldsymbol{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}}-\boldsymbol{\Gamma},\boldsymbol{\Sigma})
\end{aligned} \tag{B.15}$$

The cross terms cancel and $\mathbf{E}(\mathbf{p_k}) = \boldsymbol{\Phi}(\mathbf{C}\mathbf{x_{k|k-1}}-\boldsymbol{\Gamma},\boldsymbol{\Sigma})$ as before so that,

$$\begin{aligned}
\mathbf{R_{\tilde{y}\tilde{y}_k}} &= \mathbf{E}(\mathbf{p_k})(\mathbf{C}\mathbf{E}((\mathbf{x_k}-\mathbf{x_{k|k-1}})(\mathbf{x_k}-\mathbf{x_{k|k-1}})^{\mathbf{T}})\mathbf{C}^{\mathbf{T}}\mathbf{E}(\mathbf{p_k})^{\mathbf{T}} \\
&-\mathbf{E}(\mathbf{p_k}(\mathbf{v_k}-\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma}))(\mathbf{v_k}-\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}-\mathbf{C}\mathbf{x_{k|k-1}},\boldsymbol{\Sigma}))^{\mathbf{T}}\mathbf{p_k^T}) \\
&= \mathbf{E}(\mathbf{p_k})\mathbf{C}\boldsymbol{\Psi}_{\mathbf{k|k-1}}\mathbf{C}^{\mathbf{T}}\mathbf{E}(\mathbf{p_k})^{\mathbf{T}}+\mathbf{R_k}
\end{aligned} \tag{B.16}$$