QUANTITATIVE MEASURES OF OBSERVABILITY FOR STOCHASTIC SYSTEMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YÜKSEL SUBAŞI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

FEBRUARY 2012

Approval of the thesis:

QUANTITATIVE MEASURES OF OBSERVABILITY FOR STOCHASTIC SYSTEMS

submitted by **YÜKSEL SUBAŞI** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Department of Electrical and Electronics Engineering**, **Middle East Technical University** by,

Prof. Dr. Canan Özgen _____________
Director, Graduate School of **Natural and Applied Sciences**

Prof. Dr. İsmet Erkmen _____________
Head of Department, **Electrical and Electronics Engineering Dept.**

Prof. Dr. Mübeccel Demirekler _____________
Supervisor, **Electrical and Electronics Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Kemal Leblebicioğlu _____________
Electrical and Electronics Engineering Dept., METU

Prof. Dr. Mübeccel Demirekler _____________
Electrical and Electronics Engineering Dept., METU

Prof. Dr. Aydan Erkmen _____________
Electrical and Electronics Engineering Dept., METU

Prof. Dr. Erol Kocaoğlan _____________
Electrical and Electronics Engineering Dept., METU

Assist. Prof. Dr. Yakup Özkazanç _____________
Electrical and Electronics Engineering Dept., Hacettepe University

Date: **10.02.2012**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : YÜKSEL SUBAŞI

Signature :

ABSTRACT

QUANTITATIVE MEASURES OF OBSERVABILITY FOR STOCHASTIC SYSTEMS

Subaşı, Yüksel
Ph.D., Department of Electrical and Electronics Engineering
Supervisor: Prof. Dr. Mübeccel Demirekler

February 2012, 74 pages

The observability measure based on the mutual information between the last state and the measurement sequence originally proposed by Mohler and Hwang (1988) is analyzed in detail and improved further for linear time invariant discrete-time Gaussian stochastic systems by extending the definition to the observability measure of a state sequence. By using the new observability measure it is shown that the unobservable states of the deterministic system have no effect on this measure and any observable part with no measurement uncertainty makes it infinite. Other distance measures i.e., Bhattacharyya and Hellinger distances are also investigated to be used as observability measures.

The relationships between the observability measures and the covariance matrices of Kalman filter and the state sequence conditioned on the measurement sequence are derived. Steady state characteristics of the observability measure based on the last state is examined. The observability measures of a subspace of the state space, an individual state, the modes of the system are investigated. One of the results obtained in this part is that the deterministically unobservable states may have nonzero observability measures.

The observability measures based on the mutual information are represented recursively and calculated for nonlinear stochastic systems. Then the measures are applied to a nonlinear stochastic system by using the particle filter methods. The arguments given for the LTI case are also observed for nonlinear stochastic systems. The second moment approximation deviates from the actual values when the nonlinearity in the system increases.

Keywords: Stochastic systems, observability measure, mutual information, particle filter, entropy

ÖZ

OLASILIKSAL SİSTEMLER İÇİN GÖZLENEBİLİRLİĞİN NİCELİKSEL ÖLÇÜTLERİ

Subaşı, Yüksel
Doktora, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Mübeccel Demirekler

Şubat 2012, 74 Sayfa

Bu tez kapsamında ilk olarak Mohler ve Hwang (1988) tarafından önerilen son durum vektörü ile ölçüm dizisi arasındaki karşılıklı bilgiye dayalı gözlenebilirlik ölçütü detaylıca analiz edilmekte ve doğrusal zamandan bağımsız kesikli zamanlı Gauss gürültülü olasılıksal sistemler için son durum yerine tüm durum dizisini kullanan gözlenebilirlik ölçütü tanımına genişletilerek iyileştirilmektedir. Yeni gözlenebilirlik ölçütü kullanılarak ulaşılan sonuçlar belirli (deterministic) sistemin gözlenemeyen durumlarının gözlenebilirlik ölçütünün üzerinde etkisi olmadığını ve durum vektörünün ölçüm belirsizliği olmadan gözlenen herhangi bir kısmının ölçütü sonsuz yaptığını göstermektedir.

Tezde karşılıklı bilgi dışında diğer mesafe ölçütleri, Bhattacharyya ve Hellinger ölçütleri, de gözlenebilirlik ölçütü olarak kullanılmak amacıyla incelenmektedir.

Tez kapsamında gözlenebilirlik ölçütleriyle Kalman süzgeci değişinti matrisi ve ölçüm dizisi verildiğinde durum dizisinin değişinti matrisi arasındaki ilişkiler türetilmektedir. Tezde ayrıca son duruma dayalı gözlenebilirlik ölçütünün kararlı durum karakteristikleri incelenmektedir.

Olasılıksal sistemlerin herbir durumunun, modlarının ve durum uzayının bir alt uzayının gözlenebilirlik ölçütleri önerilen gözlenebilirlik ölçütleri kullanılarak araştırılmaktadır. Bu kısımda elde edilen sonuçlardan birisi de belirli olarak gözlenemeyen durumların sıfırdan farklı gözlenebilirlik ölçütüne sahip olabilmesidir.

Doğrusal olmayan olasılıksal sistemler için karşılıklı bilgiye dayalı gözlenebilirlik ölçütlerini kullanabilmek amacıyla bu ölçütler özyinelemeli olarak gösterilmektedir. Özyinelemeli olarak modellenen ölçütler parçacık süzgeçleri kullanılarak doğrusal olmayan sisteme uygulanmaktadır. Doğrusal sistemler için türetilen tartışmalar doğrusal olmayan olasılıksal sistemler için de gözlenmektedir. Sistemin doğrusal olmaması arttıkça, ikinci moment yaklaştırımı gerçek değerlerden uzaklaşmaktadır.

Anahtar Kelimeler: Olasılıksal sistemler, gözlenebilirlik ölçütü, karşılıklı bilgi, parçacık süzgeci, dağıntı

To my family

ACKNOWLEDGMENTS

I would like to thank all those people who have helped in the preparation of this study. I am grateful to Prof. Dr. Mübeccel Demirekler for her motivation and support that provided me determination and power throughout this study.

And, I would like to thank my family for their endless support, hope, care and help. I would like to thank my wife Tuğba Subaşı, my son İbrahim Oğuz Subaşı and my little daugther Begüm Subaşı for their patience and love and to my mother Ayşe Subaşı and my father İbrahim Subaşı for their efforts. I would like to dedicate this study to my all family whom I am proud of.

Thank you....

TABLE OF CONTENTS

ABSTRACT iv

ÖZ….. vi

ACKNOWLEDGMENTS ix

TABLE OF CONTENTS x

LIST OF TABLES xii

LIST OF FIGURES xiii

CHAPTERS

1. INTRODUCTION AND LITERATURE SURVEY 1

1.1 Motivation of the Study 4

1.2 Contributions of the Study 5

1.3 Scope of the Thesis 6

2. THEORETICAL BACKGROUND 8

2.1 Entropy 8

2.2 f-divergence 9

2.3 Kullback-Leibler Distance (Relative Entropy) 10

2.4 Mutual Information 10

2.5 Bhattacharyya Distance 11

2.6 Hellinger Distance 11

3. OBSERVABILITY MEASURES FOR LTI DISCRETE-TIME GAUSSIAN STOCHASTIC SYSTEMS 12

3.1 LTI Discrete-Time Gaussian Stochastic System 13

3.2 Observability Measures Based on the State Sequence 15

3.2.1 Observability Measure Based on the Mutual Information 15

3.2.1.1 Discussion on the Observability Measure 18

3.2.2 Observability Measure Based on the Bhattacharyya Distance 22

3.2.3 Observability Measure Based on the Hellinger Distance 24

3.3 Observability Measures Based on the Last State 25
3.3.1 Observability Measure Based on the Mutual Information....................26
3.3.1.1 Discussion on the Observability Measure..............................27
3.3.2 Observability Measure Based on the Bhattacharyya Distance 28
3.3.3 Observability Measure Based on the Hellinger Distance 29
3.4 Observability Measures for the Subspaces of the State Space 29
3.5 Observability Measure of a Single Measurement System in Observable Canonical Form..32
3.6 Discussion..36
4. COMPUTATION OF THE OBSERVABILITY MEASURES FOR NONLINEAR STOCHASTIC SYSTEMS 38
4.1 Recursive Evaluation of the Observability Measures..................................38
4.1.1 Recursive Evaluation of the Observability Measure Based on the State Sequence..39
4.1.2 Recursive Evaluation of the Observability Measure Based on the Last State..42
4.2 Particle Filter 46
4.3 Calculation of the Observability Measures by Using the Particle Filters 49
4.3.1 Calculation of the Observability Measure Based on the State Sequence49
4.3.2 Calculation of the Observability Measure Based on the Last State.......51
4.4 Computation of the Observability Measures to a Nonlinear Example 55
4.5 Discussion..62
5. CONCLUSIONS 64
REFERENCES 67
APPENDICES
A. ENTROPY OF A MULTIVARIATE NORMAL DISTRIBUTION 70
B. DETERMINANTS OF Σ_{X^k} AND $\Sigma_{[X^k,Y^k]}$...72
VITA.. 74

LIST OF TABLES

TABLES

Table 4.1 Filtering via SIS..47

Table 4.2 Resampling Algorithm..48

Table 4.3 Cross Correlation Coefficient Values...61

Table 4.4 Effects of System Parameters..62

LIST OF FIGURES

FIGURES

Figure 4.1 System State (True (Black), KF Estimate (Red), PF Estimate (Blue))........................54

Figure 4.2 $I(X^k, Y^k)$ (Theoretical (Red), PF Calculation (Blue))………………..........................54

Figure 4.3 $I(x_k, Y^k)$ (Theoretical (Red), PF Calculation (Blue))………………..........................55

Figure 4.4 Particle Filter Estimation for Case 1 (True (Black), Estimated (Red))........................57

Figure 4.5 $I(X^k, Y^k)$ for Case 1.......…………………….………………......................................58

Figure 4.6 The observability measures between the last state and the measurements: The blue line is $I(x_k, Y^k)$; Red line is $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$; Black line is the Second Moment Approximation for Case 1………...…..……….......……...................………….........................58

Figure 4.7 Correlation plot of the second moment approximation and observability measures for Case 1. Blue: $I(x_k, Y^k)$ vs. the Second Moment Approximation. Red: Difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$ vs. the Second Moment Approximation…………...58

Figure 4.8 The observability measures between the last state and the measurements: The blue line is $I(x_k, Y^k)$; Red line is $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$; Black line is the Second Moment Approximation for Case 2……….…………………..…….…………..59

Figure 4.9 Correlation plot of the second moment approximation and observability measures for Case 2. Blue: $I(x_k, Y^k)$ vs. the Second Moment Approximation. Red: Difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$ vs. the Second Moment Approximation…………..................…..........................59

Figure 4.10 The observability measures between the last state and the measurements: The blue line is $I(x_k, Y^k)$; Red line is $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$; Black line is the Second Moment Approximation for Case 3……….………………….………………...60

Figure 4.11 Correlation plot of the second moment approximation and observability measures for Case 3. Blue: $I(x_k, Y^k)$ vs. the Second Moment Approximation. Red: Difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$ vs. the Second Moment Approximation…………...60

CHAPTER 1

INTRODUCTION AND LITERATURE SURVEY

For deterministic systems, observability of a state space representation is generally handled by determining the rank conditions of the observability Gramian matrix (Kalman (1960)). The output of the process is binary; the system is either completely observable or completely unobservable. The process does not give any information about the degree of the observability. The aim of this study is to analyze the degree of observability of stochastic systems and to give an observability measure for them.

Different measures are proposed as observability measures for either deterministic or stochastic systems in the literature (Müller and Weber (1972), Koyaz (2003), Chen et al. (2007), Lindner et al. (1989), Kam et al. (1987), Mohler and Hwang (1988), Hamdan and Nayfeh (1989), Porter and Crossley (1970)). In this study, the system studied is a stochastic system. The idea is to search whether the mutual information can be used as an observability measure of this system or not; and to see the effects of any information about the initial states, process and measurement noises and system matrices on the observability measure. The mutual information is a special case of Kullback-Leibler Distance. So, other distance measures namely Bhattacharyya and Hellinger distances are also investigated to be used as observability measures.

Quantitative measures for the observability of deterministic systems are first proposed by Müller and Weber (1972). Three observability measures are presented: determinant, trace, and the maximum eigenvalue of the inverse of the observability Gramian matrix.

Inverse of the norm of the inverse of the observability Gramian is proposed as a quantitative measure of observability by Koyaz (2003). The approach depends on the generalization of some ideas presented by Müller and Weber (1972).

Mode observability is examined for a deterministic linear multivariable system which has distinct eigenvalues by Porter and Crossley (1970). The measure of the mode observability is proposed for a deterministic system described by the triple (A, B, C) when A has a set of distinct eigenvalues by Lindner et al. (1989) and Hamdan and Nayfeh (1989). The proposed measure is the angle between the rows of the matrix C and the right eigenvectors of the matrix A. The angle is not invariant under any state-coordinate transformation that is not orthogonal. By using the notion of mode observability measure proposed by Hamdan and Nayfeh (1989), time dependent measures of mode and gross observability are proposed for linear time-varying systems by Choi et al. (1999).

The observability measure, which is related to the frequency domain characteristics (zeros and residues) of a linear multivariable deterministic system, is proposed by Tarokh (1992) not only for the overall system, but also for the modes of the system. Tarokh's measure covers both distinct and repeated eigenvalues cases.

Two stochastic observability definitions, strict sense and wide sense observability, are given by Aoki (1967). The strict sense observability definition is the following. 'A stochastic system is said to be stochastically observable in the strict sense if and only if the covariance matrix associated with the conditional probability density function of the last state given all the measurements goes to zero as time goes to infinity'. The wide sense observability definition is: 'A stochastic system is said to be stochastically observable in the wide sense if the covariance matrix associated with the conditional probability density function of the last state given all the measurements remains bounded as time goes to infinity'. Definitions that are similar to the above ones are also given by Han-Fu (1980) and Bageshwar et al. (2009). However, it is known that when the system is stable and observable the conditional probability density function has a steady state covariance which may not be zero. And also when the system is stable the steady state covariance matrix of the marginal density of the states does not go to infinity (Kumar and Varaiya (1986)). The necessary condition for the stochastic observability is given as the deterministic system being observable by Han-Fu (1980) and Bageshwar et al. (2009). In addition upper and lower bounds for covariance matrix of the conditional probability density function are given by Bageshwar et al. (2009).

The degrees of observability for both the system and its subspaces are proposed by Hong et al. (2008) by using the observability Gramian for discrete linear systems.

The mutual information between the last state and the measurement sequence is proposed as the measure of the observability of a continuous-time stochastic system by Mohler and Hwang

(1988). They propose a degree of observability for a stochastic system instead of a yes or no answer for observability. The mutual information between the last state and the measurement sequence is expressed as an entropy difference between a priori and a posteriori distributions. The relationship between the mutual information and the covariance matrix associated with the conditional probability density function of the last state given all the measurements is given and it is said that this relationship holds not only for Gaussian processes, but also for some other processes of interest. They propose the second moment approximation for the nonlinear and non Gaussian stochastic systems. Another contribution of this work is to propose the same measure for the individual states. The results of Mohler and Hwang (1988) are applied to the observer path design for the bearings-only tracking by Logothetis et al. (1997). The optimal paths are derived by maximizing the mutual information between the measurement sequence and the final target state or the entire target trajectory. Dynamic programming and enumeration with optimal pruning are used as optimization techniques.

The observability measure is defined by using the mutual information and the entropy by Kam et al. (1987). The system is declared as a linear time-invariant stochastic system. However, the only probabilistic term is the discrete initial state distribution; there is no process noise or measurement noise in the model. They propose the entropy correlation coefficient

$$\rho(X, Y) = \frac{I(X, Y)}{\max\{H(X), H(Y)\}} \tag{1.1}$$

as an observability measure for a discrete state system; where, $H(.)$ is the entropy, and $I(X, Y)$ is the mutual information between the random variables X and Y.

The results of Kam et al. (1987) are extended to continuous state vector by Chen et al. (2007). They use Equation (1.1) as an observability measure for a discrete state (quantized) system. For continuous state systems, they propose to use the limits of the quantized states.

Observability and detectability are investigated for a stochastic discrete-time hybrid switching systems by West and Haddad (1994). The observability test is based on mutual information between the system states and measurements. A stochastic Popov-Belevith-Hautus Criterion for observability of stochastic systems is presented by Zhang and Chen (2004).

The observability analysis for the bearings-only tracking is considered by Le Cadre and Jauffret (1997). In the analysis, different maneuvers of target and observer are examined. The control problem is investigated for the optimization of observer maneuvers. The cost functional is based on the determinant of the Fisher information matrix.

The observability of target maneuvers via bearings-only and bearing-rate-only measurements is discussed by Hepner and Geering (1990). Intercept scenarios that result in the loss of observability are identified. A maximum likelihood estimate of the target motion is developed and analyzed for bearings-only tracking by Nardone et al. (1984). The Cramer-Rao lower bound (CRLB) is used to examine tracking strategies for bearings-only tracking by Fawcett (1988).

The problem of optimal state estimation in stochastic systems is examined by using an approach based on information theoretic measures by Feng et al. (1997). It is shown that for a linear Gaussian system, the Kalman filter is the optimal filter not only for the mean-square error measure, but for several information theoretic measures. Similar discussions are also given by Tomita et al. (1976) and Kalata and Premier (1979). These measures are minimum error-observation information, minimum error entropy, maximum mutual information between the state and the observations.

The filtering problems are studied from the viewpoint of the information theory by Tomita et al. (1976). For a linear system, it is proved that the necessary and sufficient condition for maximizing the mutual information between a state and the estimate of the state is to minimize the entropy of the estimation error.

1.1 Motivation of the Study

For the deterministic systems the observability is the ability of finding the initial state from the given outputs. Once this is done the whole state sequence can be determined using the initial state. For the stochastic systems, although the literature is quite limited, the aim is to find the role of the past measurements on the optimal estimation of the last state. Although the information gained about the last state is important in quite a large set of applications, it is also important how well we know the whole state sequence especially for the applications that does batch processing. An example from the real world may be the dim target tracking. The dim target tracking uses the complete data measured in a given time interval to find the target existence as well as the state sequence of the target. The information gained about the target’s state sequence so its observability is the ultimate aim of this problem.

Another problem that requires how the measurements related with the state sequence is the model selection problem. The representation of the system by different models will give different observability measures. A typical example of this case occurs in passive tracking where the target state is ‘unobservable’. Using modified polar coordinates to separate the unobservable part is a common practice in the literature. Model selection as in the passive tracking problem can be analyzed using the observability measure of the whole state sequence as described in this study. Another very important application area may be the sensor network design, i.e., decisions made on the number, type and placement of the sensors. The optimal path planning is another broad area that batch processing may help to design the maneuver of the own ship. The examples are not limited to the ones given above. We believe that the new observability definition given in this thesis will fill a gap that exists in this area.

Analyzing the definition of observability given in the work of Mohler and Hwang (1988) in detail to understand its behavior deeply is another motivation of this study.

The extension both of the definitions to the observability of some subspaces of the state space is considered as part of understanding the underlie structure of the observability definitions. Similarly the amount of decrease in the observability by using partial measurements is also an interesting problem that has potential applications.

The observability of nonlinear systems is another important issue. The definitions of observability do not change for the nonlinear systems however their computability is a problem. Our motivation here is to generate an algorithm to obtain numerically the observability measures of nonlinear systems. Particle filtering and Monte Carlo methods seem natural tools to solve this problem.

1.2 Contributions of the Study

The observability measure based on the mutual information between the state and the measurement sequences is proposed here for the first time. The measure is derived in terms of the statistics of the basic random variables and system matrices and analyzed in detail for LTI discrete-time Gaussian stochastic systems. Since the mutual information is a special case of Kullback-Leibler distance, other probabilistic distance measures, namely Bhattacharyya and Hellinger distances, are also investigated in full detail to be used as observability measures. The relationships between the observability measures and the covariance matrix of the state sequence conditioned on the measurement sequence are derived explicitly.

The second observability measure based on the mutual information between the last state and the measurement sequence originally proposed by Mohler and Hwang (1988) is analyzed in detail. Definition of observability of a single element of the state vector given by Mohler and Hwang is extended to observability of subspaces, so the modes, of linear Gaussian systems. Similarly it is also extended to the case of partial measurements. Partial measurement approach is unavoidable for distributed systems like sensor networks. Also Bhattacharyya and Hellinger distances can be applied to this observability measure definition.

The extension of the observability measure to the observability of a subspace of the state space is obtained for linear systems analytically. The dual of this problem is to restrict the measurements to a subspace of the measurement space. This problem can be interpreted as the sensitivity of the observability measure to the measurements. The second problem is also analyzed. The analysis of both cases given in this thesis is novel.

The observability measures based on the mutual information are expressed in a recursive manner. The recursive expressions are not only interesting in themselves but also used in the computation of observability measures of nonlinear systems. The recursions are verified with LTI discrete-time Gaussian Stochastic systems, both analytically and by simulations. The theory is applied to a simple nonlinear stochastic system by using the particle filters. The necessary relations are derived and the simulations are done. The results of the nonlinear case are compared with the linear case.

The observability measures based on the Bhattacharyya and the Hellinger distances are proposed for the first time in this thesis. In addition, the analysis given for the observability measures based on the mutual information between the last state and the measurement sequence and the mutual information between the state and the measurement sequences are new. The algorithms which are developed for the computations of the observability measures based on the mutual information for the nonlinear systems by using the particle filter and the Monte Carlo methods are new.

The observability measure analysis of the single output systems represented in the observable canonical form shed some light on understanding the ‘observability’ of the system. This analysis is done for the first time in this thesis.

1.3 Scope of the Thesis

The thesis is organized as follows:

In Chapter 1, an introduction and a literature survey is given.

In Chapter 2, the theoretical background which is needed for the other sections is included. This section summarizes shortly the concepts of entropy, mutual information, Kullback-Leibler, Bhattacharyya and Hellinger distances.

In Chapter 3 two observability measure definitions based on the measurement sequence and the state sequence or the last state are given. Observability measures used in this study are the mutual information, Bhattacharyya and Hellinger distances. These measures are explicitly derived for LTI discrete-time Gaussian stochastic systems. The results are discussed in detail. The relationships between the observability measures and Kalman filter covariance matrix and the covariance matrix of the state sequence conditioned on the measurement sequence are derived. The Observability measure of a subspace of the state space is obtained by using the same concepts and extended to an individual state and to the modes of the system. In addition, the observability measure of an individual state using the mutual information between the state and the measurement sequences is examined in detail for a single measurement system represented in observable canonical form.

In Chapter 4, the observability measures are applied to nonlinear stochastic systems by using the particle filters. In this part only the observability measures based on the mutual information are considered. First the measures are expressed in a recursive manner. Then these recursions are verified with LTI discrete-time Gaussian stochastic systems. Then necessary relations are derived to evaluate the measures by the particle filters. Also these derivations are verified with a simple LTI discrete-time Gaussian stochastic system. Finally these derivations are applied to a simple nonlinear stochastic system. The results are discussed in detail.

Chapter 5 contains the concluding remarks.

CHAPTER 2

THEORETICAL BACKGROUND

In this Chapter, the theoretical background which is needed in the later sections is given shortly. This includes the concepts of entropy, mutual information, f-divergence, Kullback-Leibler, Bhattacharyya and Hellinger distances. Detailed explanations can be found in the related references.

2.1 Entropy

Entropy is a measure of the uncertainty of a random variable (Cover and Thomas (2006)). The entropy $H(X)$ of a discrete random variable X is defined by

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{2.1}$$

where, $p(x)$ is the probability mass function of the random variable X. For a discrete random variable, the entropy is always nonnegative.

For continuous random variables, the differential entropy is defined by (Cover and Thomas (2006))

$$h(X) = -\int_S f(x) \log f(x)\, dx \tag{2.2}$$

where, S is the support set of the random variable, and $f(x)$ is the probability density function of the random variable X. Unlike discrete entropy, differential entropy may be negative. The conditional differential entropy is defined as:

$$h(X|Y) = -\iint_S f(x,y) \log f(x|y)\, dxdy = h(X,Y) - h(Y) \tag{2.3}$$

Note that the conditional entropy is the difference between the differential entropy of joint X and Y (i.e., $h(X,Y)$) and the differential entropy of Y.

2.2 f-divergence

There are many distance measures between probability distributions in the literature (Basseville (1989)). However, the distance measures used in this study are based on f-divergence.

f-divergence is a kind of measure between two probability distributions and is given in Equation (2.4), It is intuitively natural to measure the distance between the two probability densities p_1 and p_2 with the aid of the dispersion (with respect to p_1) of the likelihood ratio. More precisely, let λ be a measure on a space (X,F) such that any probability laws P_1 and P_2 are absolutely continuous with respect to λ, with densities p_1 and p_2. Let f be a continuous convex real function on R_+, and let g be an increasing function on R. Consider the following class of divergence coefficients between two probabilities:

$$d(P_1,P_2) = g\left[E_1\left[f\left(\frac{p_2}{p_1}\right)\right]\right] \tag{2.4}$$

where, $\frac{p_2}{p_1}$ is the likelihood ratio, and E_1 is the expectation with respect to P_1.

The distance measures used in the later chapters are derived from f-divergence by using the following relations.

For Kullback-Leibler distance,

$$f(x) = -\log x \qquad\qquad g(x) = x \tag{2.5}$$

$$d(P_1,P_2) = D(P_1,P_2) = \int_X p_1 \log\frac{p_1}{p_2}\, d\lambda \tag{2.6}$$

For Bhattacharyya distance,

$$f(x) = -\sqrt{x} \qquad\qquad g(x) = -\log(-x) \tag{2.7}$$

$$d(P_1,P_2) = BD(P_1,P_2) = -\log\int_X \sqrt{p_1 p_2}\, d\lambda \tag{2.8}$$

For Hellinger distance,

$$f(x) = \left(\sqrt{x}-1\right)^2 \qquad\qquad g(x) = \frac{1}{2}x \tag{2.9}$$

$$d(P_1, P_2) = HD(P_1, P_2) = \sqrt{\frac{1}{2}\int_X \left(\sqrt{p_2} - \sqrt{p_1}\right)^2 d\lambda} \quad (2.10)$$

2.3 Kullback-Leibler Distance (Relative Entropy)

Kullback-Leibler distance between two probability mass function $p(x)$ and $q(x)$ for discrete random variable X is defined by (Cover and Thomas (2006))

$$D(p, q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \quad (2.11)$$

Relative entropy is always nonnegative and is zero if and only if $p(x) = q(x)$. However, it is not a true distance between distributions, since it is not symmetric and does not satisfy the triangle inequality (Cover and Thomas (2006)).

For continuous random variables, the relative entropy between two densities $f(x)$ and $g(x)$ is defined as

$$D(f, g) = \int f(x) \log \frac{f(x)}{g(x)} dx \quad (2.12)$$

The relative entropy is always nonnegative as in the discrete case. However it is not symmetric and triangle inequality does not hold.

2.4 Mutual Information

Mutual information is a measure of the amount of information that one random variable contains about another random variable (Cover and Thomas (2006)). The mutual information between discrete random variables X and Y is defined by

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (2.13)$$

where, $p(x, y)$ is the joint probability mass function of random variables X and Y. Mutual information is a special case of the relative entropy (Kullback-Leibler (KL) distance).

The mutual information $I(X, Y)$ between the two continuous random variables with joint density $f(x, y)$ is defined as

$$I(X, Y) = \iint f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dxdy \quad (2.14)$$

The mutual information is always nonnegative, as in the discrete case.

The relationship between the entropy and the mutual information for both continuous and discrete random variables can be given as

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \qquad (2.15)$$

As can be seen from above the mutual information can be interpreted as the decrease in the entropy of a random vector due to the information obtained about it by knowing the value of the other. It can also be used as an independence measure between the two random vectors since the measure is zero only if the X and Y are independent.

2.5 Bhattacharyya Distance

Bhattacharyya Distance between the two probability density functions has the general expression (Basseville (1989)):

$$BD\left(f_x(\xi), f_y(\xi)\right) = -\log\left(BC\left(f_x(\xi), f_y(\xi)\right)\right) \quad ; \qquad 0 \le BD \le \infty \qquad (2.16)$$

where, Bhattacharyya Coefficient is defined as follows:

$$BC\left(f_x(\xi), f_y(\xi)\right) = \int \sqrt{f_x(\xi) f_y(\xi)}\, d\xi \quad ; \qquad 0 \le BC \le 1 \qquad (2.17)$$

Bhattacharyya distance is zero only if $f_x(\xi) = f_y(\xi)$, otherwise it is positive as can easily be seen from the definition. Bhattacharyya distance is symmetric, however, does not satisfy triangle inequality (Upadhyaya and Sorenson (1977)).

2.6 Hellinger Distance

Hellinger distance between two probability density functions has the general expression (Basseville (1989)):

$$HD\left(f_x(\xi), f_y(\xi)\right) = \sqrt{\frac{1}{2} \int \left(\sqrt{f_x(\xi)} - \sqrt{f_y(\xi)}\right)^2 d\xi} \quad ; \qquad 0 \le HD \le 1 \qquad (2.18)$$

Note that the Hellinger distance can be derived from the Bhattacharyya distance as:

$$HD\left(f_x(\xi), f_y(\xi)\right) = \sqrt{1 - BC\left(f_x(\xi), f_y(\xi)\right)} \qquad (2.19)$$

$$HD(P_1, P_2) = \sqrt{1 - e^{-BD(P_1, P_2)}} \qquad (2.20)$$

Hellinger distance does obey the triangle inequality (Basseville (1989) and Upadhyaya and Sorenson (1977)).

CHAPTER 3

OBSERVABILITY MEASURES FOR LTI DISCRETE-TIME GAUSSIAN STOCHASTIC SYSTEMS

In this section two observability measure definitions are given. The first definition is the observability measure of the state sequence $\{x_i\}_{i=0}^{k}$. This definition is not a very familiar one however in some problems it is more meaningful to obtain some information about the whole state sequence rather than the initial state x_0 or the last state x_k. The second definition is a more traditional one and is defined as the information obtained about the last state. The observability measures used in this section are the mutual information, Bhattacharyya and Hellinger distances.

A short summary of the section is as follows:

The observability measures, which are based on the mutual information, Bhattacharyya and Hellinger distances, are derived for LTI discrete-time Gaussian stochastic systems. The derivations are given explicitly. The results are discussed in detail.

The relationships between the observability measures (the last state– measurement sequence observability or the state sequence-measurement sequence observability) and Kalman filter covariance matrix and the covariance matrix of the state sequence conditioned on the measurement sequence are derived. Steady state characteristics of the observability measure based on the last state is examined. An individual state, a mode and a subspace of the state space observability measures of stochastic systems are investigated by using the proposed observability measures.

A single measurement system represented in observable canonical form is examined in detail and observability measure is obtained for the individual state sequences. One of the results obtained in this part is that the deterministically unobservable states have nonzero observability measures.

3.1 LTI Discrete-Time Gaussian Stochastic System

For a LTI discrete-time Gaussian stochastic system, the system equations are given as

$$x_{k+1} = Ax_k + Gw_k \tag{3.1}$$

$$y_k = Cx_k + Hv_k \tag{3.2}$$

where, $x_k \in R^n$, $y_k \in R^m$ is the measurement of the system. A, G, C, H are constant matrices. It is assumed that $\{x_0, w_k \in R^p, v_k \in R^r\}_{k=0}^{\infty}$ are independent, $\{w_k\}_{k=0}^{\infty}$ and $\{v_k\}_{k=0}^{\infty}$ are both identically distributed and

$$x_0 \sim N(\overline{x_0}, \Sigma_0); \quad w_k \sim N(0, Q); \quad v_k \sim N(0, R) \tag{3.3}$$

Definition 3.1: x_0, $\{w_k\}$ and $\{v_k\}$ are defined as the basic random variables.

For time k, the state equation of the system can be written in terms of the basic random variables as:

$$x_k = A^k x_0 + \sum_{i=0}^{k-1} A^{k-1-i} G w_i \tag{3.4}$$

which yields the measurement equation:

$$y_k = CA^k x_0 + C\sum_{i=0}^{k-1} A^{k-1-i} G w_i + Hv_k \tag{3.5}$$

Since the aim is to obtain an observability measure for the complete state sequence the above expressions are defined in a more compact form. By using above Equations (3.4) and (3.5), the following equations can be written.

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} I \\ A \\ A^2 \\ \vdots \\ A^k \end{bmatrix} x_0 + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ G & 0 & \cdots & 0 \\ AG & G & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A^{k-1}G & A^{k-2}G & \cdots & G \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{k-1} \end{bmatrix} \tag{3.6}$$

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^k \end{bmatrix} x_0 + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ CG & 0 & \cdots & 0 \\ CAG & CG & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{k-1}G & CA^{k-2}G & \cdots & CG \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{k-1} \end{bmatrix} + \begin{bmatrix} H & 0 & \cdots & 0 \\ 0 & H & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_k \end{bmatrix} \tag{3.7}$$

These equations are linear equations of the state and the output sequences in terms of the basic random variables. For notational simplicity we will define the following random vectors and matrices.

$$X^k = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}_{(k+1)n \times 1} ; Y^k = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}_{(k+1)m \times 1} ; A_k = \begin{bmatrix} I \\ A \\ A^2 \\ \vdots \\ A^k \end{bmatrix}_{(k+1)n \times n} \tag{3.8}$$

$$C_k = \begin{bmatrix} C & 0 & \cdots & 0 \\ 0 & C & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \end{bmatrix}_{(k+1)m \times (k+1)n} ; \; H_k = \begin{bmatrix} H & 0 & \cdots & 0 \\ 0 & H & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H \end{bmatrix}_{(k+1)m \times (k+1)r} \tag{3.9}$$

$$G_k = \begin{bmatrix} 0 & 0 & \cdots & & 0 \\ G & 0 & \cdots & & 0 \\ AG & G & \ddots & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ A^{k-1}G & A^{k-2}G & \cdots & & G \end{bmatrix}_{(k+1)n \times kp} ; Q_k = \begin{bmatrix} Q & 0 & \cdots & 0 \\ 0 & Q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q \end{bmatrix}_{kp \times kp} \tag{3.10}$$

$$R_k = \begin{bmatrix} R & 0 & \cdots & 0 \\ 0 & R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R \end{bmatrix}_{(k+1)r \times (k+1)r} ; \; W^k = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{k-1} \end{bmatrix}_{kp \times 1} ; \; V^k = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_k \end{bmatrix}_{(k+1)r \times 1} \tag{3.11}$$

Now one can write the dynamic equation and the measurement equation as:

$$X^k = A_k x_0 + G_k W^k \tag{3.12}$$

$$Y^k = C_k A_k x_0 + C_k G_k W^k + H_k V^k \tag{3.13}$$

Theorem 3.1: The random vectors X^k and Y^k have normal densities, their means are $A_k\overline{x_0}$ and $C_k A_k \overline{x_0}$. Their covariance matrices are given by:

$$\Sigma_{X^k} = A_k \Sigma_0 {A_k}^T + G_k Q_k {G_k}^T \tag{3.14}$$

$$\Sigma_{Y^k} = C_k A_k \Sigma_0 {A_k}^T {C_k}^T + C_k G_k Q_k {G_k}^T {C_k}^T + H_k R_k {H_k}^T = C_k \Sigma_{X^k} {C_k}^T + H_k R_k {H_k}^T \tag{3.15}$$

$$\Sigma_{X^k Y^k} = \Sigma_{X^k} {C_k}^T \tag{3.16}$$

Proof: Gaussianity is trivial. The covariance matrices of X^k and Y^k can be easily obtained from Equations (3.12) and (3.13) by considering the independence of the basic random variables. The last equality can be obtained by considering the covariance matrix of the joint random vectors X^k and Y^k

$$\Sigma_{[X^k, Y^k]} = \begin{bmatrix} \Sigma_{X^k} & \Sigma_{X^k Y^k} \\ \Sigma_{Y^k X^k} & \Sigma_{Y^k} \end{bmatrix} \tag{3.17}$$

where the cross covariance matrix of X^k and Y^k is

$$\Sigma_{X^k Y^k} = E\left(\left(X^k - E\left(X^k\right)\right)\left(Y^k - E\left(Y^k\right)\right)^T\right) \tag{3.18}$$

$$\Sigma_{X^k Y^k} = E\left(\left(A_k(x_0 - \overline{x_0}) + G_k W^k\right)\left(C_k A_k (x_0 - \overline{x_0}) + C_k G_k W^k + H_k V^k\right)^T\right) \tag{3.19}$$

$$\Sigma_{X^kY^k} = A_k\Sigma_0A_k^{\ T}C_k^{\ T} + G_kQ_kG_k^{\ T}C_k^{\ T} = \Sigma_{X^k}C_k^{\ T} \tag{3.20}$$

■

Theorem 3.1 gives the necessary matrices that are used in the observability measures.

3.2 Observability Measures Based on the State Sequence

The observability measures are defined as distances between the complete state sequence X^k and the complete set of measurements Y^k. In this section, the definitions of the observability measures based on the observability of the complete state sequence are given for LTI discrete-time Gaussian stochastic systems.

3.2.1 Observability Measure Based on the Mutual Information

Definition 3.2: The observability measure is the mutual information between the state sequence X^k and the measurement sequence Y^k.

The following theorem gives a simple expression for the observability measure for the linear Gaussian system described above. A special case of this theorem is given in Huang, T., Chen, B. (2008)).

Theorem 3.2: The variables X^k and Y^k have Gaussian distributions and the mutual information between these variables can be calculated as

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X^k}||\Sigma_{Y^k}|}{|\Sigma_{[X^k,Y^k]}|} \tag{3.21}$$

where $|.|$ is the determinant.

Proof: By using Equation (2.14), the mutual information between X^k and Y^k can be calculated as:

$$I(X^k, Y^k) = \iint f(X^k, Y^k)\log\frac{f(X^k, Y^k)}{f(X^k)f(Y^k)}dX^kdY^k \tag{3.22}$$

where,

$$f(X^k, Y^k) = \frac{1}{\sqrt{(2\pi)^{(n+m)(k+1)}|\Sigma_{[X^k,Y^k]}|}}e^{-\frac{1}{2}\left(\left(X^k-E(X^k)\right)^T\ \left(Y^k-E(Y^k)\right)^T\right)\Sigma_{[X^k,Y^k]}{}^{-1}\begin{pmatrix}\left(X^k-E(X^k)\right)\\ \left(Y^k-E(Y^k)\right)\end{pmatrix}} \tag{3.23}$$

$$f(X^k) = \frac{1}{\sqrt{(2\pi)^{n(k+1)}|\Sigma_{X^k}|}} e^{-\frac{1}{2}\left(X^k - E(X^k)\right)^T \Sigma_{X^k}{}^{-1}\left(X^k - E(X^k)\right)} \tag{3.24}$$

$$f(Y^k) = \frac{1}{\sqrt{(2\pi)^{m(k+1)}|\Sigma_{Y^k}|}} e^{-\frac{1}{2}\left(Y^k - E(Y^k)\right)^T \Sigma_{Y^k}{}^{-1}\left(Y^k - E(Y^k)\right)} \tag{3.25}$$

By substituting Equations (3.23), (3.24) and (3.25) into Equation (3.22) the mutual information between X^k and Y^k can be found as:

$$I(X^k, Y^k) = \iint f(X^k, Y^k)\left(\log f(X^k, Y^k) - \log f(X^k) - \log f(Y^k)\right) dX^k dY^k \tag{3.26}$$

$$\begin{aligned} I(X^k, Y^k) = & \iint f(X^k, Y^k)\log f(X^k, Y^k) dX^k dY^k - \int f(X^k)\log f(X^k) dX^k \\ & - \int f(Y^k)\log f(Y^k) dY^k \end{aligned} \tag{3.27}$$

$$I(X^k, Y^k) = -h(X^k, Y^k) + h(X^k) + h(Y^k) \tag{3.28}$$

By using the derivation given in Appendix-A one can write:

$$\begin{aligned} I(X^k, Y^k) = & -\frac{(n+m)(k+1)}{2} - \frac{1}{2}\log(2\pi)^{(n+m)(k+1)}\left|\Sigma_{[X^k,Y^k]}\right| + \frac{n(k+1)}{2} \\ & + \frac{1}{2}\log(2\pi)^{n(k+1)}\left|\Sigma_{X^k}\right| + \frac{m(k+1)}{2} + \frac{1}{2}\log(2\pi)^{m(k+1)}\left|\Sigma_{Y^k}\right| \end{aligned} \tag{3.29}$$

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{(2\pi)^{n(k+1)}(2\pi)^{m(k+1)}\left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k}\right|}{(2\pi)^{(n+m)(k+1)}\left|\Sigma_{[X^k,Y^k]}\right|} \tag{3.30}$$

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{\left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k}\right|}{\left|\Sigma_{[X^k,Y^k]}\right|} \tag{3.31}$$

■

The above theorem gives the mutual information between the state and the measurements in terms of the covariance matrices of the related variables. The aim of the next theorem is to obtain the same mutual information in terms of the basic system parameters.

Theorem 3.3: The mutual information between X^k and Y^k is

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{\left|C_k A_k \Sigma_0 A_k{}^T C_k{}^T + C_k G_k Q_k G_k{}^T C_k{}^T + H_k R_k H_k{}^T\right|}{\left|H_k R_k H_k{}^T\right|} \tag{3.32}$$

Proof: Determinant of $\Sigma_{[X^k,Y^k]}$ can be written as:

$$\left|\Sigma_{[X^k,Y^k]}\right| = \begin{vmatrix} \Sigma_{X^k} & \Sigma_{X^kY^k} \\ \Sigma_{Y^kX^k} & \Sigma_{Y^k} \end{vmatrix} = \left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k} - \Sigma_{Y^kX^k}\Sigma_{X^k}{}^{-1}\Sigma_{X^kY^k}\right| \tag{3.33}$$

By substituting Equations (3.15) and (3.16) into Equation (3.33),

$$\left|\Sigma_{[X^k,Y^k]}\right| = \left|\Sigma_{X^k}\right|\left|C_k\Sigma_{X^k}C_k{}^T + H_kR_kH_k{}^T - C_k\Sigma_{X^k}\Sigma_{X^k}{}^{-1}\Sigma_{X^k}C_k{}^T\right| \tag{3.34}$$

$$\left|\Sigma_{[X^k,Y^k]}\right| = \left|\Sigma_{X^k}\right|\left|C_k\Sigma_{X^k}C_k{}^T - C_k\Sigma_{X^k}C_k{}^T + H_kR_kH_k{}^T\right| = \left|\Sigma_{X^k}\right|\left|H_kR_kH_k{}^T\right| \tag{3.35}$$

Note that

$$\left|H_kR_kH_k{}^T\right| = \begin{vmatrix} HRH^T & 0 & \cdots & 0 \\ 0 & HRH^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & HRH^T \end{vmatrix} = \left|HRH^T\right|^{k+1} \tag{3.36}$$

Substitution of Equation (3.35) into Equation (3.31) gives the observability measure as:

$$\begin{aligned} I(X^k,Y^k) &= \frac{1}{2}\log\frac{\left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k}\right|}{\left|\Sigma_{X^k}\right|\left|H_kR_kH_k{}^T\right|} \\ &= \frac{1}{2}\log\frac{\left|C_kA_k\Sigma_0A_k{}^TC_k{}^T + C_kG_kQ_kG_k{}^TC_k{}^T + H_kR_kH_k{}^T\right|}{\left|H_kR_kH_k{}^T\right|} \end{aligned} \tag{3.37}$$

■

Theorem 3.3 gives the mutual information in terms of the system matrices. Some conclusions about observability will be obtained from this result. However some other convenient forms of this expression that are useful will be given first.

Fact 3.2.1: The observability measure is

$$I(X^k,Y^k) = \frac{1}{2}\log\frac{\left|\Sigma_0\right|\left|GQG^T\right|^k}{\left|\Sigma_{X^k} - \Sigma_{X^kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kX^k}\right|} \tag{3.38}$$

Proof: By using the derivation ($\left|\Sigma_{X^k}\right|$ and $\left|\Sigma_{[X^k,Y^k]}\right|$) given in Appendix-B, Equation (3.37) can be written as

$$I(X^k,Y^k) = \frac{1}{2}\log\frac{\left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k}\right|}{\left|\Sigma_{Y^k}\right|\left|\Sigma_{X^k} - \Sigma_{X^kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kX^k}\right|} = \frac{1}{2}\log\frac{\left|\Sigma_0\right|\left|GQG^T\right|^k}{\left|\Sigma_{X^k} - \Sigma_{X^kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kX^k}\right|} \tag{3.39}$$

■

Fact 3.2.2: The observability measure is

$$I(X^k,Y^k) = \frac{1}{2}\log\frac{\left|\Sigma_{X^k}\right|}{\left|\Sigma_{X^k|Y^k}\right|} \tag{3.40}$$

Proof: The denominator term in Equation (3.39) is equal to $\Sigma_{X^k|Y^k}$ which is the covariance matrix of the state sequence conditioned on the measurement sequence (Logothetis et al. (1997)).

■

Several conclusions can be derived from the above given results to understand the meaning of it as well as to assess the measure.

3.2.1.1 Discussion on the Observability Measure

One can make several observations to criticize the proposed observability measure. The observations are listed in below as facts.

Before giving the list one should point out that in the derivation of the observability measure given in Equation (3.32), the term $\left|\Sigma_{X^k}\right|$ is cancelled. When $|\Sigma_0|$ and/or $\left|GQG^T\right|$ are singular this creates a problem that the mutual information becomes indefinite. For this case, the limit of $I\left(X^k, Y^k\right)$ as $|\Sigma_0|$ and/or $\left|GQG^T\right|$ approaches to zero can be considered as the actual measure.

Note also that the observability measure compares the determinants of the conditional and unconditional covariance matrices of the state sequence.

Fact 3.2.1.1.1: Unobservable states of the pair (C, A) have no effect on the observability measure.

Proof: Let (C, A) be written in the following form:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \qquad C = \begin{bmatrix} 0 & C_{12} \end{bmatrix} \tag{3.41}$$

where A_{11} contains the unobservable modes of (C, A). Assume that Σ_0 and GQG^T are partitioned accordingly:

$$\Sigma_0 = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad GQG^T = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \tag{3.42}$$

The terms of the mutual information given in Equation (3.32) can be written in terms of the basic system parameters as follows:

$$C_k A_k \Sigma_0 {A_k}^T {C_k}^T = \begin{bmatrix} C\Sigma_0 C^T & C\Sigma_0 A^T C^T & \cdots & C\Sigma_0 {A^k}^T C^T \\ CA\Sigma_0 C^T & CA\Sigma_0 A^T C^T & \cdots & CA\Sigma_0 {A^k}^T C^T \\ \vdots & \vdots & \ddots & \vdots \\ CA^k\Sigma_0 C^T & CA^k\Sigma_0 A^T C^T & \cdots & CA^k\Sigma_0 {A^k}^T C^T \end{bmatrix} \tag{3.43}$$

$$C_k G_k Q_k {G_k}^T {C_k}^T = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & CGQG^T C^T & \cdots & CGQG^T {A^{k-1}}^T C^T \\ \vdots & \vdots & \ddots & \vdots \\ 0 & CA^{k-1}GQG^T C^T & \cdots & C\left(\sum_{i=0}^{k-1} A^{k-1-i} GQG^T {A^{k-1-i}}^T\right) C^T \end{bmatrix} \tag{3.44}$$

If we use the matrices given in Equations (3.41) and (3.42) the following result can be found.

$$CA^r\Sigma_0 A^{p^T}C^T = C_{12}A_{22}^r\Sigma_{22}A_{22}^{p\ T}C_{12}^{\ T}; \qquad p = 0, \dots, k; \qquad r = 0, \dots, k \tag{3.45}$$

So the term $C_k A_k \sum_0 A_k^{\ T} C_k^{\ T}$ is independent of the unobservable part. Similarly $C_k G_k Q_k G_k^{\ T} C_k^{\ T}$ contains the terms of $C_{12}A_{22}^r q_{22} A_{22}^{p\ T} C_{12}^{\ T}$. So again the unobservable part is not involved. ■

The above fact seems to be paradoxical in the sense that the observability measure is not zero when we have unobservable states. This is obviously not the case because of two reasons, one is that the measurements provide us some information about the state sequence and the second is that the classical observability of the deterministic sates is concentrated on the computation of the initial state. However, even if the initial state cannot be determined, its later values may be found by using the available measurements. Later we will be talking about the 'amount of observability' of unobservable states to make this point more clear.

Fact 3.2.1.1.2: When the observation noise is zero, the observability measure is equal to infinity. Even if there is only one noiseless measurement i.e., HRH^T is singular, this fact holds.

Proof: Trivial from Equations (3.32) and (3.36). ■

This property suggests the following treatments:

a. The observability measure gives infinity because of the existence of the outputs with no noise since they provide a perfect estimation for some part of the state vector. So we change our aim as obtaining an observability measure for the remaining states by eliminating the perfectly estimated part. The procedure is described as an algorithm below.

   i. Define $\bar{y}_k \triangleq My_k$ so that $MHv_k = \begin{bmatrix} H_1 \\ 0 \end{bmatrix} v_k = \begin{bmatrix} v_1 \\ 0 \end{bmatrix}$, where M is nonsingular.

   ii. Define a transformation on the state of the system as $\tilde{x}_k = Tx_k$ so that $\tilde{C} = MCT^{-1} = \begin{bmatrix} \tilde{C}_{11} & \tilde{C}_{12} \\ I & 0 \end{bmatrix}$, where I is the identity matrix and T is nonsingular. The new system equations are

$$\begin{bmatrix} \tilde{x}_{k+1}^1 \\ \tilde{x}_{k+1}^2 \end{bmatrix} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} \tilde{x}_k^1 \\ \tilde{x}_k^2 \end{bmatrix} + \begin{bmatrix} \tilde{w}_k^1 \\ \tilde{w}_k^2 \end{bmatrix} \tag{3.46}$$

$$\begin{bmatrix} \tilde{y}_k^1 \\ \tilde{y}_k^2 \end{bmatrix} = \begin{bmatrix} \tilde{C}_{11} & \tilde{C}_{12} \\ I & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_k^1 \\ \tilde{x}_k^2 \end{bmatrix} + \begin{bmatrix} v_1 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} \tilde{y}_k^1 - \tilde{C}_{11}\tilde{y}_k^2 \\ \tilde{y}_k^2 \end{bmatrix} = \begin{bmatrix} 0 & \tilde{C}_{12} \\ I & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_k^1 \\ \tilde{x}_k^2 \end{bmatrix} + \begin{bmatrix} v_1 \\ 0 \end{bmatrix} \tag{3.47}$$

$\tilde{x}_k^1$ is measured directly with $\tilde{y}_k^2$. The state equation of $\tilde{x}_k^2$ is

$$\tilde{x}_{k+1}^2 = \tilde{A}_{21}\tilde{x}_k^1 + \tilde{A}_{22}\tilde{x}_k^2 + \tilde{w}_k^2 \tag{3.48}$$

Note that $\tilde{A}_{21}\tilde{x}_k^1$ is known and it can be treated as a known input of the system. Then the system is reduced to

$$\tilde{x}_{k+1}^2 = \tilde{A}_{22}\tilde{x}_k^2 + u_k + \tilde{w}_k^2; \qquad \hat{y}_k = \tilde{C}_{12}\tilde{x}_k^2 + v_1 \tag{3.49}$$

where $\hat{y}_k = \tilde{y}_k^1 - \tilde{C}_{11}\tilde{y}_k^2$ and $u_k = \tilde{A}_{21}\tilde{x}_k^1$. As a result, the proposed observability measure can be applied to the reduced system.

a. Another approach to overcome the perfect measurement problem is to add a fixed small noise to all measurements. This approach helps to differentiate the observability of two perfectly measured modes. After this treatment they may have different observability indices.

Fact 3.2.1.1.3: The observability measure is bounded from below as $I(X^k, Y^k) = \frac{1}{2}\log\left|I_{n(k+1)} + MM^T\right| \geq \frac{1}{2}\log(1 + |M|^2)$ where $|M|^2 \propto \left|\sqrt{\Sigma_{X^k}}\right|$.

Proof: Let us make a nonsingular transformation on the measurement equation such that $HRH^T = I_m$ where I_m is the mxm identity matrix. Then the observability measure becomes $I(X^k, Y^k) = \frac{1}{2}\log\left|I_{m(k+1)} + \tilde{C}_k\Sigma_{X^k}\tilde{C}_k{}^T\right|$ where $\tilde{C}$ is the new measurement matrix after the transformation. By using the determinant property $|I_m + A_{mxn}B_{nxm}| = |I_n + B_{nxm}A_{mxn}|$ the measure becomes $I(X^k, Y^k) = \frac{1}{2}\log\left|I_{n(k+1)} + \tilde{C}_k{}^T\tilde{C}_k\Sigma_{X^k}\right|$. Note that $\tilde{C}_k{}^T\tilde{C}_k$ and Σ_{X^k} are both positive definite for which the square roots exist. Let $S = \sqrt{\tilde{C}_k^T\tilde{C}_k}$ and $T = \sqrt{\Sigma_{X^k}}$ .Using the determinant property again, the measure will be $I(X^k, Y^k) = \frac{1}{2}\log\left|I_{n(k+1)} + STTS\right|$. Define $M \triangleq ST$, then the measure will be $I(X^k, Y^k) = \frac{1}{2}\log\left|I_{n(k+1)} + MM^T\right|$. Note that $|M| = \left|\sqrt{\tilde{C}_k^T\tilde{C}_k}\right|\left|\sqrt{\Sigma_{X^k}}\right|$. If we use the inequality, $|S + V| \geq |S| + |V|$ for positive definite matrices, we obtain $\frac{1}{2}\log\left|I_{n(k+1)} + MM^T\right| \geq \frac{1}{2}\log(1 + |M|^2)$. ■

The above inequality shows the relationship between the 'largeness' of several system matrices and the mutual information. It also indicates that when the determinant of the state sequence covariance matrix Σ_{X^k} increases the lower bound of the observability measure increases.

Fact 3.2.1.1.4: If the determinant of the covariance matrix of the process noise, Q, increases, the observability measure increases.

Proof: Trivial from Equation (3.37) ■

Fact 3.2.1.1.5: If the determinant of the initial covariance matrix, Σ_0, increases, the observability measure increases.

Proof: Trivial from Equation (3.37). ■

Fact 3.2.1.1.6:. When there are no process noise and the initial state uncertainty, the observability measure is zero.

Proof: Trivial from Equation (3.37). ■

The above given facts shed further light on the definition of the observability measure as the mutual information. The facts can be interpreted as that the observability measure is actually a measure for the value of the measurement in the determination of the state vector. If the uncertainty of the state is large then the measurement is more valuable so one gets a larger value as the observability measure. Fact 3.2.1.1.6 tells that the measurements are not valuable at all when the initial uncertainty and the process noise are zero since the initial state is known exactly and the system has no uncertainty.

The Facts 3.2.1.1.7 and 3.2.1.1.8 are intuitive; they are similar to our conception of observability. They simply tell that if the measurement noise increases than the observability measure decreases.

Fact 3.2.1.1.7: If the determinant of the matrix, C^TC, increases the lower bound of the observability measure increases.

Proof: Trivial from the proof of the Fact 3.2.1.1.3. ■

Fact 3.2.1.1.8: If the determinant of the covariance matrix of the measurement noise, R, increases, the lower bound of the observability measure decreases.

Proof:

$$\frac{\left|C_k\Sigma_{X^k}{C_k}^T + H_kR_k{H_k}^T\right|}{\left|H_kR_k{H_k}^T\right|} = \frac{\left|H_kR_k{H_k}^T\right|\left|\left(H_kR_k{H_k}^T\right)^{-1}C_k\Sigma_{X^k}{C_k}^T + I_{m(k+1)}\right|}{\left|H_kR_k{H_k}^T\right|} \tag{3.50}$$

$$\frac{\left|C_k\Sigma_{X^k}{C_k}^T + H_kR_k{H_k}^T\right|}{\left|H_kR_k{H_k}^T\right|} = \left|\left(H_kR_k{H_k}^T\right)^{-1}C_k\Sigma_{X^k}{C_k}^T + I_{m(k+1)}\right| \tag{3.51}$$

By using the same procedure given in the proof of the Fact 3.2.1.1.3, it is seen that the determinant of the covariance matrix of the measurement noise increases, the lower bound of the observability measure decreases. Note that $\left|H_kR_k{H_k}^T\right| = \left|HRH^T\right|^{k+1}$. ■

Defining observability measure as the mutual information between the measurements and the states seems to be natural. Considering the fact that the mutual information is a special case of the Kullback-Leibler distance, it can be stated that the observability measure reduces to the Kullback-Leibler distance between the two Gaussians. Based on this fact, it is thought that other distance measures may also be used to define an observability measure for stochastic systems. In the following subsection, we will derive Equations for the Bhattacharyya distance (BD) and the Hellinger distance (HD) based observability measures by using the state and the measurement sequences.

3.2.2 Observability Measure Based on Bhattacharyya Distance

Definition 3.3: The observability measure is the Bhattacharyya distance between the two Gaussian densities $f(X^k, Y^k)$ and $f(X^k)f(Y^k)$.

The covariance matrix of the joint probability density function $f(X^k, Y^k)$ of X^k and Y^k is given in Equation (3.17), and the density function of $f(X^k)f(Y^k)$ is,

$$f(X^k)f(Y^k) = \frac{1}{\sqrt{(2\pi)^{(n+m)(k+1)}\left|\Sigma_{(X^k.Y^k)}\right|}} e^{-\frac{1}{2}\left(\left(X^k - E(X^k)\right)^T \left(Y^k - E(Y^k)\right)^T\right)\Sigma_{(X^k.Y^k)}{}^{-1}\begin{pmatrix}\left(X^k - E(X^k)\right)\\ \left(Y^k - E(Y^k)\right)\end{pmatrix}} \tag{3.52}$$

where, $\Sigma_{(X^k.Y^k)} = \begin{bmatrix}\Sigma_{X^k} & 0\\ 0 & \Sigma_{Y^k}\end{bmatrix}$.

Notice that, the only difference between two Gaussian density functions $f(X^k, Y^k)$ and $f(X^k)f(Y^k)$ is the covariance matrices. The covariance matrix of $f(X^k)f(Y^k)$ does not involve cross covariance matrices $\Sigma_{Y^kX^k}$ and $\Sigma_{X^kY^k}$ of $f(X^k, Y^k)$. These equations are used in the following theorems.

Theorem 3.4: Bhattacharyya distance between $f(X^k, Y^k)$ and $f(X^k)f(Y^k)$ is

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}C_k\Sigma_{X^k}C_k{}^T + H_kR_kH_k{}^T\right|}{\left|HRH^T\right|^{\frac{k+1}{2}}\left|\Sigma_{Y^k}\right|^{\frac{1}{2}}} \tag{3.53}$$

Proof: Bhattacharyya distance between these two Gaussian densities can be found from the following equation (Fukunaga (1990)).

$$\begin{aligned} &BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) \\ &= \frac{1}{8}\left(\begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix} - \begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix}\right)^T \left(\frac{\Sigma_{[X^k,Y^k]} + \Sigma_{(X^k.Y^k)}}{2}\right)^{-1}\left(\begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix} \right. \\ &\left. - \begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix}\right) + \frac{1}{2}\log\frac{\left|\frac{\Sigma_{[X^k,Y^k]} + \Sigma_{(X^k.Y^k)}}{2}\right|}{\sqrt{\left|\Sigma_{[X^k,Y^k]}\right|\left|\Sigma_{(X^k.Y^k)}\right|}} \end{aligned} \tag{3.54}$$

The first term in the right side of Equation (3.54) is zero since the two densities have the same mean value. Then

$$\left|\frac{\Sigma_{[X^k,Y^k]} + \Sigma_{(X^k.Y^k)}}{2}\right| = \left|\begin{bmatrix} \Sigma_{X^k} & \frac{\Sigma_{X^kY^k}}{2} \\ \frac{\Sigma_{Y^kX^k}}{2} & \Sigma_{Y^k} \end{bmatrix}\right| = \left|\Sigma_{X^k}\right|\left|\frac{3}{4}C_k\Sigma_{X^k}C_k{}^T + H_kR_kH_k{}^T\right| \tag{3.55}$$

and

$$\left|\Sigma_{[X^k,Y^k]}\right|\left|\Sigma_{(X^k.Y^k)}\right| = \left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k} - \Sigma_{Y^kX^k}\Sigma_{X^k}{}^{-1}\Sigma_{X^kY^k}\right|\left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k}\right| \tag{3.56}$$

$$\left|\Sigma_{[X^k,Y^k]}\right|\left|\Sigma_{(X^k.Y^k)}\right| = \left|\Sigma_{X^k}\right|^2\left|\Sigma_{Y^k}\right|\left|HRH^T\right|^{k+1} \tag{3.57}$$

By substituting Equations (3.55) and (3.57) into Equation (3.54) the observability measure can be found as

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}C_k\Sigma_{X^k}C_k{}^T + H_kR_kH_k{}^T\right|}{\left|HRH^T\right|^{\frac{k+1}{2}}\left|\Sigma_{Y^k}\right|^{\frac{1}{2}}} \tag{3.58}$$

■

By examining Equation (3.58), it can be seen easily that the discussions given in Section 3.2.1.1 are also valid for this measure. Unobservable states of the pair (C, A) have also no effect on this observability measure since Equation (3.58) includes the same matrices. When $\left|HRH^T\right|$ is singular the measure is equal to infinity. Note that the numerator term in Equation (3.58) is equal to $\left|\Sigma_{Y^k} - \frac{1}{4} H_k R_k {H_k}^T\right|$ and is very similar to the numerator term in Equation (3.32). Therefore when the determinants of the initial state covariance Σ_0, the covariance matrix of the process noise Q, the state sequence covariance matrix Σ_{X^k} and the matrix $C^T C$, increases, the measure increases as in the mutual information case since the power of the denominator term of $\left|\Sigma_{Y^k}\right|$ is 0.5. Also when the determinants of the covariance matrix of the measurement noise increases, the observability measure decreases. It is seen easily from Equation (3.58) that when there is no process noise and the initial state uncertainty, the observability measure is zero.

Fact 3.2.2.1: The observability measure can be written in terms of the conditional and unconditional covariance matrices of the state sequence as:

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}\Sigma_{X^k} + \frac{1}{4}\Sigma_{X^k|Y^k}\right|}{\sqrt{\left|\Sigma_{X^k}\right|\left|\Sigma_{X^k|Y^k}\right|}} \tag{3.59}$$

Proof: By using Equation (3.54) the following relation can be found:

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\Sigma_{Y^k}\right|\left|\Sigma_{X^k} - \frac{1}{4}\Sigma_{X^kY^k}{\Sigma_{Y^k}}^{-1}\Sigma_{Y^kX^k}\right|}{\sqrt{\left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k}\right|\left|\Sigma_{Y^k}\right|\left|\Sigma_{X^k} - \Sigma_{X^kY^k}{\Sigma_{Y^k}}^{-1}\Sigma_{Y^kX^k}\right|}} \tag{3.60}$$

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}\Sigma_{X^k} + \frac{1}{4}\Sigma_{X^k|Y^k}\right|}{\sqrt{\left|\Sigma_{X^k}\right|\left|\Sigma_{X^k|Y^k}\right|}} \tag{3.61}$$

■

3.2.3 Observability Measure Based on Hellinger Distance

Definition 3.4: The observability measure is the Hellinger distance between $f(X^k, Y^k)$ and $f(X^k)f(Y^k)$.

Theorem 3.5: The Hellinger distance between $f(X^k, Y^k)$ and $f(X^k)f(Y^k)$ is

$$HD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \sqrt{1 - \frac{\left|HRH^T\right|^{\frac{k+1}{4}}\left|\Sigma_{Y^k}\right|^{\frac{1}{4}}}{\left|\frac{3}{4}C_k\Sigma_{X^k}{C_k}^T + H_k R_k {H_k}^T\right|^{\frac{1}{2}}}} \tag{3.62}$$

Proof: The relation can be found easily from Equations (3.58) and (2.20). ■

Equation (3.62) implies again that the discussions given for the mutual information case and the Bhattacharyya case are also valid for this measure. Note that the maximum value of the Hellinger distance is one.

Fact 3.2.3.1: The observability measure can be written in terms of the conditional and unconditional covariance matrices of the state sequences as:

$$\mathrm{HD}\left(f\left(X^k, Y^k\right), f\left(X^k\right)f\left(Y^k\right)\right) = \sqrt{1 - \frac{\left(\left|\Sigma_{X^k}\right|\left|\Sigma_{X^k|Y^k}\right|\right)^{\frac{1}{4}}}{\left|\frac{3}{4}\Sigma_{X^k} + \frac{1}{4}\Sigma_{X^k|Y^k}\right|^{\frac{1}{2}}}} \tag{3.63}$$

Proof: The relation can be easily found from Equations (3.59) and (2.20). ■

The mutual information between the states and the outputs for time k can be found by using the same procedure as:

$$I(x_k, y_k) = \frac{1}{2}\log\left|C\Sigma_{x_k}C^T + HRH^T\right| - \frac{1}{2}\log\left|HRH^T\right| \tag{3.64}$$

where Σ_{x_k} is the covariance matrix of x_k. Note that, when HRH^T is singular, the mutual information goes to infinity.

By using the same procedure for the two Gaussian densities $f(x_k, y_k)$ and $f(x_k)f(y_k)$ the Bhattacharyya Distance can be found as:

$$\mathrm{BD}\left(f(x_k, y_k), f(x_k)f(y_k)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}C\Sigma_{x_k}C^T + HRH^T\right|}{\left|HRH^T\right|^{\frac{1}{2}}\left|\Sigma_{y_k}\right|^{\frac{1}{2}}} \tag{3.65}$$

Also Hellinger distance between $f(x_k, y_k)$ and $f(x_k)f(y_k)$ can be found by using Equations (3.65) and (2.20) as:

$$\mathrm{HD}\left(f(x_k, y_k)\ , f(x_k)f(y_k)\right) = \sqrt{1 - \frac{\left|HRH^T\right|^{\frac{1}{4}}\left|\Sigma_{y_k}\right|^{\frac{1}{4}}}{\left|\frac{3}{4}C\Sigma_{x_k}C^T + HRH^T\right|^{\frac{1}{2}}}} \tag{3.66}$$

3.3 Observability Measures Based on the Last State

In some applications knowledge about the last state may be important, so the definition of the observability measure can be changed to concentrate on the last state. This definition is the one given in Mohler and Hwang (1988). In this section we will give further analysis of this observability measure.

In Mohler and Hwang (1988) the observability measure is defined for continuous time stochastic systems. The consequences of the definition are not exploited however the definition is applied to 'the bearings only tracking' problem. In this part we exploit the definition almost fully for linear time invariant discrete time Gaussian stochastic systems, extend the definition to the Bhattacharyya and the Hellinger distances. Also the analysis given below is firstly presented for this measure in this thesis.

3.3.1 Observability Measure Based on the Mutual Information

Definition 3.5: The observability measure is the mutual information between the last state x_k and the measurement sequence Y^k (Mohler and Hwang (1988)).

The following theorem that gives the computation of this measure is from Mohler and Hwang (1988).

Theorem 3.6: The observability measure is related with the ratio between the determinant of the marginal density covariance matrix Σ_{x_k} of x_k and the determinant of the Kalman filter state covariance matrix $\Sigma_{x_k|Y^k}$ at time k. That is

$$I(x_k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{x_k}|}{|\Sigma_{x_k|Y^k}|} \tag{3.67}$$

Proof: : Kalman Filter state covariance at time k is (Anderson and Moore (1979)),

$$\Sigma_{x_k|Y^k} = \Sigma_{x_k} - \Sigma_{x_k Y^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^k x_k} \tag{3.68}$$

where, $\Sigma_{x_k|Y^k}$ is the conditional covariance matrix of x_k given all measurements Y^k at time k, Σ_{x_k} is the covariance matrix of x_k, Σ_{Y^k} is the covariance matrix of Y^k, $\Sigma_{x_k Y^k}$ and $\Sigma_{Y^k x_k}$ are cross covariances matrices of x_k and Y^k.

The mutual information between x_k and Y^k can be written as given in Equation (3.21)

$$I(x_k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{x_k}||\Sigma_{Y^k}|}{|\Sigma_{[x_k,Y^k]}|} \tag{3.69}$$

where, Σ_{x_k,Y^k} is covariance matrix of joint probability density function of x_k and Y^k,

$$\Sigma_{[x_k,Y^k]} = \begin{bmatrix} \Sigma_{x_k} & \Sigma_{x_k Y^k} \\ \Sigma_{Y^k x_k} & \Sigma_{Y^k} \end{bmatrix} \tag{3.70}$$

Determinant of $\Sigma_{[x_k,Y^k]}$ is,

$$\left|\Sigma_{[x_k,Y^k]}\right| = \left|\Sigma_{Y^k}\right|\left|\Sigma_{x_k} - \Sigma_{x_kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kx_k}\right| \tag{3.71}$$

Note that, the second term on the right side of Equation (3.71) is the determinant of the Kalman Filter state covariance at time k. By substituting Equation (3.71) into Equation (3.69), the mutual information between x_k and Y^k can be found as:

$$I(x_k, Y^k) = \frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|\left|\Sigma_{Y^k}\right|}{\left|\Sigma_{Y^k}\right|\left|\Sigma_{x_k} - \Sigma_{x_kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kx_k}\right|} \tag{3.72}$$

$$I(x_k, Y^k) = \frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|}{\left|\Sigma_{x_k} - \Sigma_{x_kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kx_k}\right|} \tag{3.73}$$

$$I(x_k, Y^k) = \frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|}{\left|\Sigma_{x_k|Y^k}\right|} \tag{3.74}$$

■

The state covariance matrix of the Kalman filter certainly represents the uncertainty of the state at any time k so it may be argued that this can also be used as an observability measure.

3.3.1.1 Discussion on the Observability Measure

The second observability measure which is based on the last state has similar properties as the first measure which is based on the state sequence. They are summarized below.

Fact 3.3.1.1.1: When HRH^T is singular, the observability measure is equal to infinity.

Proof: When HRH^T is singular, $\left|\Sigma_{x_k|Y^k}\right|$ will be zero and the measure will be equal to infinity. ■

To overcome this difficulty, the procedure given in 3.2.1.1 can be used.

Fact 3.3.1.1.2: When the system is stable, the initial uncertainty on the state does not affect the observability measure at the steady state.

Proof: The numerator given in Equation (3.67) is the determinant of $\Sigma_{x_k} = A^k\Sigma_0 A^{k^T} + \sum_{i=0}^{k-1} A^{k-1-i}GQG^T A^{k-1-i^T}$ which is independent of the initial covariance when $k \to \infty$ if the system is stable. The denominator is the determinant of the conditional covariance so again independent of the initial covariance. ■

Fact 3.3.1.1.3: When there are no process noise and no initial state uncertainty, the observability measure is zero.

Proof: System is deterministic and the initial state is known so the states can be found exactly and the measurements will not give any new information. Therefore $\Sigma_{x_k|Y^k}$ is equal to Σ_{x_k} and the measure is zero. ■

Fact 3.3.1.1.4: When the covariance matrix of the measurement noise decreases, the observability measure increases.

Proof: Note that the observability measure is inversely proportional with the conditional covariance of the state. By looking at the Kalman filter equations it is seen that the conditional covariance of the state decreases with decreasing the measurement noise. ■

Fact 3.3.1.1.5: If the determinant of the matrix, $C^T C$, increases the observability measure increases.

Proof: Trivial from the fact that increasing the determinant of the matrix $C^T C$ corresponds to a decrease in the covariance of the measurement noise. ■

3.3.2 Observability Measure Based on Bhattacharyya Distance

Definition 3.6: The observability measure is the Bhattacharyya distance between the two Gaussian densities $f(x_k, Y^k)$ and $f(x_k)f(Y^k)$.

Theorem 3.7: The Bhattacharyya distance between the two Gaussian densities $f(x_k, Y^k)$ and $f(x_k)f(Y^k)$ is

$$BD\left(f(x_k, Y^k), f(x_k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}\Sigma_{x_k} + \frac{1}{4}\Sigma_{x_k|Y^k}\right|}{\sqrt{\left|\Sigma_{x_k}\right|\left|\Sigma_{x_k|Y^k}\right|}} \tag{3.75}$$

Proof: By using Equation (3.54) the following relations can be written.

$$BD\left(f(x_k, Y^k), f(x_k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\Sigma_{Y^k}\right|\left|\Sigma_{x_k} - \frac{1}{4}\Sigma_{x_kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kx_k}\right|}{\sqrt{\left|\Sigma_{x_k}\right|\left|\Sigma_{Y^k}\right|\left|\Sigma_{Y^k}\right|\left|\Sigma_{x_k} - \Sigma_{x_kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kx_k}\right|}} \tag{3.76}$$

$$\mathrm{BD}\left(f\left(x_k, Y^k\right), f(x_k)f\left(Y^k\right)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}\Sigma_{x_k} + \frac{1}{4}\Sigma_{x_k|Y^k}\right|}{\sqrt{\left|\Sigma_{x_k}\right|\left|\Sigma_{x_k|Y^k}\right|}} \tag{3.77}$$

■

By looking at Equation (3.77) it is seen that the discussions given in 3.3.1.1 can also be applied to this measure.

3.3.3 Observability Measure Based on Hellinger Distance

Definition 3.7: The observability measure is the Hellinger distance between the two Gaussian densities $f\left(x_k, Y^k\right)$ and $f(x_k)f\left(Y^k\right)$.

Theorem 3.8: The Hellinger distance between the two Gaussian densities $f\left(x_k, Y^k\right)$ and $f(x_k)f\left(Y^k\right)$ is

$$\mathrm{HD}\left(f\left(x_k, Y^k\right), f(x_k)f\left(Y^k\right)\right) = \sqrt{1 - \frac{\left(\left|\Sigma_{x_k}\right|\left|\Sigma_{x_k|Y^k}\right|\right)^{\frac{1}{4}}}{\left|\frac{3}{4}\Sigma_{x_k} + \frac{1}{4}\Sigma_{x_k|Y^k}\right|^{\frac{1}{2}}}} \tag{3.78}$$

Proof: The relation can be found easily from Equations (3.77) and (2.20). ■

By looking at Equation (3.78) it is seen that the discussions given in 3.3.1.1 can also be applied to this measure.

3.4 Observability Measures for the Subspaces of the State Space

In this section the observability measure of a subspace of the state space is analyzed. We use the mutual information definition of the observability measure in this section. Both the sequence and the final state definitions are given. In particular, it is started by analyzing the observability measure of a one dimensional subspace that corresponds to an element of the state vector. For individual states of the stochastic system, the mutual information between an individual state and the observations can be used to define the observability measure.

Definition 3.8: The observability measure of a subspace of the state space represented as $\gamma_k = Mx_k$ of a stochastic system is defined as the mutual information between γ_k and the measurement sequence that is, $I\left(\gamma_k, Y^k\right)$.

From the previous analysis we can compute $\mathrm{I}(\gamma_k, Y^k)$ as:

$$\mathrm{I}(\gamma_k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{\gamma_k}|}{|\Sigma_{\gamma_k|Y^k}|} \tag{3.79}$$

Note that the matrix M can be chosen such that only a particular state variable is selected as γ_k by choosing an appropriate row for M. In Mohler and Hwang (1988) only this special case is introduced with no analysis.

The definition and the related result given in Equation (3.79) can be extended to the observability measure of a subspace sequence $\{\gamma_i\}_{i=0}^{k}$ by using Equation (3.40). Define $\Gamma^k = \{\gamma_i\}_{i=0}^{k}$.

Definition 3.9: The observability measure of Γ^k is defined as the mutual information between Γ^k and the measurement sequence.

Similar arguments as given above leads writing the mutual information in terms of the covariance matrices as follows:

$$\mathrm{I}(\Gamma^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{\Gamma^k}|}{|\Sigma_{\Gamma^k|Y^k}|} \tag{3.80}$$

One can also define observability measures of γ_k and Γ^k by considering only a part of the measurement sequence.

Definition 3.10: The observability measures of γ_k and Γ^k by considering only a part of the measurement sequence is defined as the mutual information between γ_k or Γ^k and a part of the measurement sequence Y_j^k. That is,

$$\mathrm{I}(\gamma_k, Y_j^k) = \frac{1}{2}\log\frac{|\Sigma_{\gamma_k}|}{|\Sigma_{\gamma_k|Y_j^k}|} \tag{3.81}$$

$$\mathrm{I}(\Gamma^k, Y_j^k) = \frac{1}{2}\log\frac{|\Sigma_{\Gamma^k}|}{|\Sigma_{\Gamma^k|Y_j^k}|} \tag{3.82}$$

When the state equation is in Jordan form, these equations can be used to define the observability measures of the system modes. Also similar derivations can be done for the observability measures which use Bhattacharyya and Hellinger distances.

Example 3.1: We will apply the above ideas to a simple example. Let the system equations be defined as:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} -0.5 & 0 \\ 0 & -0.7 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + w_k \tag{3.83}$$

$$y_k = \begin{bmatrix} 0.75 & 0.075 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + v_k \tag{3.84}$$

where

$$x_0 \sim N\left(\overline{x_0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), w_k \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\right), v_k \sim N(0,0.5) \tag{3.85}$$

Note that the system representation is selected in the diagonal form so that the observability measures of the individual states become the observability measures of the modes. The process noise is selected so that both states are under the process noises that are independent but having same statistics. Since the coefficient of the first state in the matrix C is higher than the coefficient of the second state, the expectation is that the first state is more observable compared to the second one. The computed observability measures of the states $x_{1,k}$ and $x_{2,k}$ according to Equation (3.79) are: $I(x_{1,k}, Y^k) = 0.3127$ and $I(x_{2,k}, Y^k) = 0.0044$. These values are for the steady state i.e., a large k and satisfy the expectation. Similarly for the state sequence, the computed observability measures of X_1^k and X_2^k according to Equation (3.80) are $I(X_1^k, Y^k) = 25.4810$ and $I(X_2^k, Y^k) = 0.2383$, as expected, for $k = 100$. And also the observability measures of x_k and X^k according to Equations (3.67) and (3.32) are $I(x_k, Y^k) = 0.3206$ and $I(X^k, Y^k) = 26.0296$.

If we change the matrix A to $\begin{bmatrix} -0.1 & 0 \\ 0 & -0.9 \end{bmatrix}$, the results are; $I(x_k, Y^k) = 0.2678$, $I(x_{1,k}, Y^k) = 0.2206$, $I(x_{2,k}, Y^k) = 0.0416$, $I(X^k, Y^k) = 23.6108$, $I(X_1^k, Y^k) = 22.3159$, $I(X_2^k, Y^k) = 0.8035$. As seen from the results, $I(x_k, Y^k)$ and $I(X^k, Y^k)$ values decrease. This is actually comes from the fact that the changes of the determinant of the matrix A decreases with respect to the above example. Even if the changes of the determinant of the matrix A do not affect the determinant of the Σ_{X^k}, the eigenvalues of Σ_{X^k} change so that the observability measure $I(X^k, Y^k)$ decrease.

$I(x_{1,k}, Y^k)$ and $I(X_1^k, Y^k)$ decreases more as compared with $I(x_k, Y^k)$ and $I(X^k, Y^k)$. This is because, while the eigenvalue of this mode decreases the eigenvalue of the second mode increases. In addition, since the eigenvalue of the second mode increases $I(x_{2,k}, Y^k)$ and $I(X_2^k, Y^k)$ values increase.

3.5 Observability Measure of a Single Measurement System in Observable Canonical Form

In this part, the observability measure based on the mutual information for a single measurement system which is represented in observable canonical form is analyzed in detail. The relationships between the observability measures for the system and the individual state sequences are examined.

Let the system equations be defined in observable canonical form as:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \vdots \\ x_{n,k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & a_1 \\ 1 & 0 & 0 & a_2 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 1 & a_n \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{bmatrix} + Gw_k \tag{3.86}$$

$$y_k = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{bmatrix} + Hv_k \tag{3.87}$$

By using the above system representation we will analyze the variation of the observability measure among the states.

Theorem 3.9: When the system equations are given as in Equations (3.86) and (3.87), the observability measure of the system $I(X^k, Y^k)$ is equal to the observability measure of the n^th state $I(X_n^k, Y^k)$, i.e., $I(X_n^k, Y^k) = I(X^k, Y^k)$.

Proof: The observability measure of the system $I(X^k, Y^k)$ is given in Equation (3.32) as:

$$\begin{aligned} I(X^k, Y^k) &= \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{|H_k R_k H_k{}^T|} \\ &= \frac{1}{2}\log\frac{|C_k A_k \Sigma_0 A_k{}^T C_k{}^T + C_k G_k Q_k G_k{}^T C_k{}^T + H_k R_k H_k{}^T|}{|H_k R_k H_k{}^T|} \end{aligned} \tag{3.88}$$

It is easy to see that both the numerator and the denominator of this expression are the related covariance of the last state because of the special structure of the matrix C. The covariance matrix of the measurement sequence is given in Equation (3.15). By using the measurement matrix C given in Equation (3.87), the covariance matrix of the measurement sequence is

$$\Sigma_{Y^k} = \Sigma_{X_n^k} + H_k R_k H_k{}^T \tag{3.89}$$

Again by using the measurement matrix properties, the cross covariance of X_n^k and the measurement sequence is

$$\Sigma_{X_n^k Y^k} = \Sigma_{X_n^k} \tag{3.90}$$

The observability measure of the nth state $I(X_n^k, Y^k)$ can be calculated as

$$I(X_n^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X_n^k}||\Sigma_{Y^k}|}{|\Sigma_{[X_n^k, Y^k]}|} = \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{|\Sigma_{Y^k} - \Sigma_{Y^k X_n^k}\Sigma_{X_n^k}{}^{-1}\Sigma_{X_n^k Y^k}|} \tag{3.91}$$

And by using Equations (3.89) and (3.90), the following relation can be found.

$$\Sigma_{Y^k} - \Sigma_{Y^k X_n^k}\Sigma_{X_n^k}{}^{-1}\Sigma_{X_n^k Y^k} = \Sigma_{X_n^k} + H_k R_k H_k{}^T - \Sigma_{X_n^k}\Sigma_{X_n^k}{}^{-1}\Sigma_{X_n^k} = H_k R_k H_k{}^T \tag{3.92}$$

By using Equations (3.91) and (3.92), the observability measure of the nth state can be found as

$$I(X_n^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{|H_k R_k H_k{}^T|} = I(X^k, Y^k) \tag{3.93}$$

■

Theorem 3.10: The observability measure of the ith state $I(X_i^k, Y^k)$ $(i \neq n)$ is smaller than or equal to the observability measure of the nth state $I(X^k, Y^k)$, i.e., $I(X_i^k, Y^k) \leq I(X_n^k, Y^k)$.

Proof: By using Equation (3.91) the observability measure of the ith state $I(X_i^k, Y^k)$ is

$$I(X_i^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{|\Sigma_{Y^k} - \Sigma_{Y^k X_i^k}\Sigma_{X_i^k}{}^{-1}\Sigma_{X_i^k Y^k}|} \tag{3.94}$$

The cross covariance of X_i^k and Y^k is

$$\Sigma_{X_i^k Y^k} = \Sigma_{X_i^k X_n^k} \tag{3.95}$$

By substituting Equations (3.89) and (3.95) into Equation (3.94), one can get

$$I(X_i^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{|\Sigma_{X_n^k} + H_k R_k H_k{}^T - \Sigma_{X_n^k X_i^k}\Sigma_{X_i^k}{}^{-1}\Sigma_{X_i^k X_n^k}|} \tag{3.96}$$

$$I(X_i^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{|H_k R_k H_k{}^T + \Sigma_{X_n^k} - \Sigma_{X_n^k X_i^k}\Sigma_{X_i^k}{}^{-1}\Sigma_{X_i^k X_n^k}|} \tag{3.97}$$

The term $\Sigma_{X_n^k} - \Sigma_{X_n^k X_i^k}\Sigma_{X_i^k}{}^{-1}\Sigma_{X_i^k X_n^k}$ is the covariance matrix of $X_n^k | X_i^k$. When this covariance matrix is equal to zero, the observability measure of the ith state $I(X_i^k, Y^k)$ becomes equal to the observability measure of the system $I(X^k, Y^k)$. Otherwise it is smaller than the observability measure of the system $I(X^k, Y^k)$ because the denominator term in Equation (3.97) increases.

■

Theorem 3.11: When the system is given in observable canonical form as in Equations (3.86) and (3.87), if there is no process noise (i.e., w_k is a zero vector), the observability measure of the i^th state $I(X_i^k, Y^k)$ $(i = 1, \dots, n)$ is equal to the observability measure of the system $I(X^k, Y^k)$ for all i.

$$I(X_i^k, Y^k) = I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{|H_k R_k H_k{}^T|}; \qquad (i = 1, \dots, n); \tag{3.98}$$

Proof: Since the system is given in the observable canonical form, if the one of the states is known exactly, other states can be calculated exactly in n steps from this state information. When this is the case, the term $\Sigma_{X_n^k} - \Sigma_{X_n^k X_i^k}\Sigma_{X_i^k}{}^{-1}\Sigma_{X_i^k X_n^k}$ in Equation (3.97) becomes equal to the zero matrix and $I(X_i^k, Y^k)$ becomes equal to $I(X^k, Y^k)$. ■

Note that Theorems 3.9-3.11 use only the measurement matrix C properties of the above system representation, they are also valid for the systems having the same measurement matrix.

Theorem 3.12: Unobservable states of the pair (C, A) of the system given in Equations (3.1) and (3.2) have zero observability measure when the off-block diagonal terms of Σ_0 and GQG^T are zero and when A_{12} given in Equation (3.41) is a zero matrix, otherwise they have nonzero observability measure.

Proof: Let us assume that the system matrices are given as

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \qquad C = \begin{bmatrix} 0 & C_{12} \end{bmatrix} \tag{3.99}$$

$$\Sigma_0 = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \qquad GQG^T = \begin{bmatrix} q_{11} & 0 \\ 0 & q_{22} \end{bmatrix} \tag{3.100}$$

for the system given in Equations (3.1) and (3.2). It is seen in the measurement matrix that X_2^k is measured and X_1^k is not measured and unobservable. The observability measure of X_1^k can be found from the following relation by using Equation (3.97)

$$I(X_1^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{Y^k}|}{\left|\Sigma_{Y^k} - \Sigma_{X_2^k X_1^k}\Sigma_{X_1^k}{}^{-1}\Sigma_{X_1^k X_2^k}\right|} \tag{3.101}$$

The observability measure is zero only when $\Sigma_{X_1^k X_2^k}$ is zero. $\Sigma_{X_1^k X_2^k}$ is composed of the terms $\left\{\Sigma_{x_{1,i} x_{2,j}}\right\}_{i=0,j=0}^{k,k}$.

Let us use the system matrices given in Equations (3.99) and (3.100). Σ_{X^k} contains the following terms:

$$A^r\Sigma_0 A^{p^T} = \begin{bmatrix} A_{11}^r\Sigma_{11}A_{11}^{p^T} & 0 \\ 0 & A_{22}^r\Sigma_{22}A_{22}^{p^T} \end{bmatrix}; \qquad p = 0,\dots,k; \qquad r = 0,\dots,k \tag{3.102}$$

and

$$A^r GQG^T A^{p^T} = \begin{bmatrix} A_{11}^r GQG^T A_{11}^{p^T} & 0 \\ 0 & A_{22}^r GQG^T A_{22}^{p^T} \end{bmatrix}; \quad p = 0,\dots,k; \quad r = 0,\dots,k \tag{3.103}$$

As seen from Equations (3.102) and (3.103) that when the off-block diagonal terms of Σ_0 and GQG^T are zero and when A_{12} given in Equation (3.41) is a zero matrix, the terms $\left\{\Sigma_{x_{1,i}x_{2,j}}\right\}_{i=0,j=0}^{k,k}$ (upper right in the matrices given in Equations (3.102) and (3.103)) are zero so $\Sigma_{X_1^k X_2^k}$ is a zero matrix. Otherwise it is not zero and the observability measures of the unobservable states are nonzero. ■

The results of this theorem can be seen in Example 3.3.

Example 3.2: Here we apply these ideas to an example. Let the system equations be defined in the observable canonical form as:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ x_{3,k+1} \\ x_{4,k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0.2 \\ 1 & 0 & 0 & 0.3 \\ 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0.1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \\ x_{4,k} \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \tag{3.104}$$

$$y_k = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \\ x_{4,k} \end{bmatrix} + v_k \tag{3.105}$$

where,

$$x_0 \sim N\left(\overline{x_0}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\right), w_k \sim N\left(0, \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}\right), v_k \sim N(0,0.5) \tag{3.106}$$

For $k = 100$ the observability measures can be found as: $I(X^k, Y^k) = 84.2120$, $I(X_1^k, Y^k) = 18.6910$, $I(X_2^k, Y^k) = 31.7573$, $I(X_3^k, Y^k) = 52.0732$, $I(X_4^k, Y^k) = 84.2120$. Note that $I(X^k, Y^k)$ and $I(X_4^k, Y^k)$ are equal to each other. And also for $i = 1,\dots,3$ the mutual information between X_i^k and Y^k $I(X_i^k, Y^k)$ is less than $I(X^k, Y^k)$.

If we use the same system equations and if there is no process noise, the observability measures can be found as 6.6182 for $I(X^k, Y^k)$ and $I(X_i^k, Y^k)$ $(i = 1,\dots,4)$. The observability measure obtained is much smaller than the observability measures obtained for the non-zero process noise case. This fact can be explained by investigating the related covariance matrices. However one can also explain the phenomena by arguing that ‘if the process noise is zero then the only uncertainty in the state is due to the uncertainty of the initial state and the observations provide information only to the value of the initial state’.

Example 3.3: In this example we will see an interesting result which shows that even if the individual state is unobservable, the observability measure of that individual state may not be zero. Let the system equations be defined as:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ x_{3,k+1} \\ x_{4,k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 0.3 \\ 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0.1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \\ x_{4,k} \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \tag{3.107}$$

$$y_k = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \\ x_{4,k} \end{bmatrix} + v_k \tag{3.108}$$

where,

$$x_0 \sim N\left(\overline{x_0}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right), w_k \sim N\left(0, \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix} \right), v_k \sim N(0,0.5) \tag{3.109}$$

Note that the first state is not observable, (A, C) pair is not full column rank (rank=3). For $k = 100$, the observability measures can be found as: $I\left(X^k, Y^k\right) = 73.2754$, $I\left(X_1^k, Y^k\right) = 7.7544$, $I\left(X_2^k, Y^k\right) = 22.3305$, $I\left(X_3^k, Y^k\right) = 42.0457$, $I\left(X_4^k, Y^k\right) = 73.2754$.

Note that again $I\left(X^k, Y^k\right)$ and $I\left(X_4^k, Y^k\right)$ are equal to each other. Even if the first state is unobservable, the observability measure of the state is not zero. This comes from the fact that the uncertainty of the unobservable state decreases by using the information coming from the estimation of the observable states. Also if the state which drives the unobservable state is more observable, the observability measure of the unobservable state increase. In addition if the unobservable state is also driven by the unobservable states, the observability measure decreases because of the increasing in the uncertainty which is not observable.

3.6 Discussion

The observability measures based on the mutual information between the state and the measurement sequences and the mutual information between the last state and the measurement sequence are derived explicitly and analyzed in detail. The definitions are extended by using other probabilistic distance measures which are the Bhattacharyya and the Hellinger distances.

By using the observability measure based on the mutual information between the state and the measurement sequences it is shown that the unobservable states of the deterministic system have no effect on this measure. In spite of this fact the observability measures of the unobservable states considered individually may not be zero.

Another important observation is that any observable part with no measurement uncertainty makes the measure infinite. This is reasonable in the sense that some part of the state is known exactly which corresponds to infinite information.

Both of the observability measure definitions are extended to the observability measure of any subspace of the state space. Another important extension is the observability measure obtained by using partial measurements. Partial measurement approach is unavoidable for distributed systems like sensor networks. Also the Bhattacharyya and the Hellinger distances can be applied to these observability measure definitions.

The analysis of several observability measures defined in this section shows that they do not contribute much to the observability measure concept over the basic mutual information definition. However the Hellinger distance may be preferable because of its boundedness. The definition is clearly expandable to non Gaussian and nonlinear systems.

CHAPTER 4

COMPUTATION OF THE OBSERVABILITY MEASURES FOR NONLINEAR STOCHASTIC SYSTEMS

In this Chapter, the observability measures based on the mutual information between the measurement and the state sequences; and the mutual information between the measurement sequence and the last state are calculated for nonlinear stochastic systems. The measures are represented recursively to be applied to nonlinear stochastic systems. The measures are applied to a nonlinear stochastic system by using the particle filter methods. Finally these derivations are applied to a simple nonlinear stochastic system. The results are discussed in detail.

4.1 Recursive Evaluation of the Observability Measures

In this section the two observability measure definitions are represented recursively. They are

- the mutual information between the state and the measurement sequences
- the mutual information between the last state and the measurement sequence

The recursive evaluation of the measures is itself an asset but at the same time it is an unavoidable requirement for the computation of the observability measures for the nonlinear systems. In the recursions, the only assumption is that the basic random variables of the system are independent and the process noise and the measurement noise sequences are identically distributed.

For a discrete-time stochastic system, the system equations that are used in this study can be given as:

$$x_{k+1} = p(x_k, w_k) \tag{4.1}$$

$$y_k = g(x_k, v_k) \tag{4.2}$$

where $x_k \in R^n$, $y_k \in R^m$ are the state and the measurement of the system. It is assumed that, the basic random variables $\{x_0, w_k, v_k\}_{k=0}^{\infty}$ are all independent and $\{w_k\}_{k=0}^{\infty}$ and $\{v_k\}_{k=0}^{\infty}$ are both identically distributed.

4.1.1 Recursive Evaluation of the Observability Measure Based on the State Sequence

In this subsection the observability measure which is the mutual information between the state and the measurement sequences $I(X^k, Y^k)$ are expressed in a recursive manner.

The state sequence X^k and the measurement sequence Y^k are represented as

$$X^k = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}; \qquad Y^k = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \tag{4.3}$$

The mutual information between X^k and Y^k can be calculated as

$$I(X^k, Y^k) = \iint f(X^k, Y^k) \log \frac{f(X^k, Y^k)}{f(X^k) f(Y^k)} dX^k dY^k \tag{4.4}$$

where, $f(X^k)$ and $f(Y^k)$ are the probability density functions of X^k and Y^k respectively, $f(X^k, Y^k)$ is the joint probability density function of X^k and Y^k.

Theorem 4.1: The observability measure based on the mutual information between the state and the measurement sequences can be written recursively as:

$$I(X^k, Y^k) = I(X^{k-1}, Y^{k-1}) - h(y_k|x_k) + h(y_k|Y^{k-1}) \tag{4.5}$$

Proof: $f(X^k, Y^k)$ can be written as:

$$f(X^k, Y^k) = f(y_k|X^k, Y^{k-1}) f(x_k|X^{k-1}, Y^{k-1}) f(X^{k-1}, Y^{k-1}) \tag{4.6}$$

By assuming the basic random variables are independent and the process noise and the measurement noise are identically distributed, this relation can be written as:

$$f(X^k, Y^k) = f(y_k|x_k) f(x_k|x_{k-1}) f(X^{k-1}, Y^{k-1}) \tag{4.7}$$

Same assumptions lead also

$$f(X^k) = f(x_k|x_{k-1}) f(X^{k-1}) \tag{4.8}$$

$$f(Y^k) = f(y_k|Y^{k-1}) f(Y^{k-1}) \tag{4.9}$$

By substituting Equations (4.7), (4.8) and (4.9) into Equation (4.4), one can get

$$I(X^k, Y^k) = \iint f(X^k, Y^k) \log \frac{f(y_k|x_k) f(x_k|x_{k-1}) f(X^{k-1}, Y^{k-1})}{f(x_k|x_{k-1}) f(X^{k-1}) f(y_k|Y^{k-1}) f(Y^{k-1})} dX^k dY^k \tag{4.10}$$

$$I(X^k, Y^k) = \iint f(X^k, Y^k) \left(\log \frac{f(X^{k-1}, Y^{k-1})}{f(X^{k-1}) f(Y^{k-1})} + \log \frac{f(y_k|x_k)}{f(y_k|Y^{k-1})} \right) dX^k dY^k \tag{4.11}$$

$$\begin{aligned} I(X^k, Y^k) = \iint f(X^k, Y^k) \log \frac{f(X^{k-1}, Y^{k-1})}{f(X^{k-1}) f(Y^{k-1})} dX^k dY^k \\ + \iint f(X^k, Y^k) \log \frac{f(y_k|x_k)}{f(y_k|Y^{k-1})} dX^k dY^k \end{aligned} \tag{4.12}$$

The integration of the first term on the right side of Equation (4.12) first with respect to y_k and then x_k, gives us the following.

$$\begin{aligned} \iint f(X^k, Y^k) \log \frac{f(X^{k-1}, Y^{k-1})}{f(X^{k-1}) f(Y^{k-1})} dX^k dY^k \\ = \iint f(X^{k-1}, Y^{k-1}) \log \frac{f(X^{k-1}, Y^{k-1})}{f(X^{k-1}) f(Y^{k-1})} dX^{k-1} dY^{k-1} \end{aligned} \tag{4.13}$$

which is equal to $I(X^{k-1}, Y^{k-1})$.

Integrating the second term in Equation (4.12) with respect to appropriate variables we obtain the following.

$$\begin{aligned} \iint f(X^k, Y^k) \log \frac{f(y_k|x_k)}{f(y_k|Y^{k-1})} dX^k dY^k &= \iint f(X^k, Y^k) \log f(y_k|x_k)\, dX^k dY^k \\ &- \iint f(X^k, Y^k) \log f(y_k|Y^{k-1})\, dX^k dY^k \\ &= \iint f(x_k, y_k) \log f(y_k|x_k) dx_k dy_k - \int f(Y^k) \log f(y_k|Y^{k-1}) dY^k \end{aligned} \tag{4.14}$$

The final equation can be written in terms of the conditional entropies.

$$\iint f(X^k, Y^k) \log \frac{f(y_k|x_k)}{f(y_k|Y^{k-1})} dX^k dY^k = -h(y_k|x_k) + h(y_k|Y^{k-1}) \tag{4.15}$$

By substituting Equations (4.13) and (4.15) into Equation (4.12), one can get

$$I(X^k, Y^k) = I(X^{k-1}, Y^{k-1}) - h(y_k|x_k) + h(y_k|Y^{k-1}) \tag{4.16}$$

■

Equation (4.16) indicates that when the information gained from x_k about y_k is more valuable than the information gained from Y^{k-1} about y_k the mutual information increases in time.

Equation (4.5) is not recursive in a strict sense since the term $h(y_k|Y^{k-1})$ depends on the past values of the measurements. We will suggest a method to overcome this problem later.

To elaborate the recursive mutual entropy concept further similar recursions for linear time invariant systems are obtained and compared with Equation (4.16).

The mutual information between the state and the measurement sequences for an LTI discrete-time stochastic system is

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X^k}||\Sigma_{Y^k}|}{|\Sigma_{[X^k,Y^k]}|} \tag{4.17}$$

Note that

$$|\Sigma_{X^k}| = |\Sigma_{X^{k-1}}||GQG^T| \tag{4.18}$$

$$|\Sigma_{[X^k,Y^k]}| = |\Sigma_{[X^{k-1},Y^{k-1}]}|\,|GQG^T||HRH^T| \tag{4.19}$$

$$f(Y^k) = f(y_k|Y^{k-1})f(Y^{k-1}) \tag{4.20}$$

$$|\Sigma_{Y^k}| = |\Sigma_{Y^{k-1}}|\,|\Sigma_{y_k|Y^{k-1}}| \tag{4.21}$$

$$f(y_k, x_k) = f(y_k|x_k)f(x_k) \tag{4.22}$$

$$|\Sigma_{[x_k,y_k]}| = |\Sigma_{y_k|x_k}||\Sigma_{x_k}| \tag{4.23}$$

By substituting these relations into Equation (4.17) one can get

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X^{k-1}}||GQG^T||\Sigma_{Y^{k-1}}|\,|\Sigma_{y_k|Y^{k-1}}|}{|\Sigma_{[X^{k-1},Y^{k-1}]}||GQG^T||HRH^T|} = \frac{1}{2}\log\frac{|\Sigma_{X^{k-1}}||\Sigma_{Y^{k-1}}|\,|\Sigma_{y_k|Y^{k-1}}|}{|\Sigma_{[X^{k-1},Y^{k-1}]}||HRH^T|} \tag{4.24}$$

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X^{k-1}}||\Sigma_{Y^{k-1}}|}{|\Sigma_{[X^{k-1},Y^{k-1}]}|} + \frac{1}{2}\log|\Sigma_{y_k|Y^{k-1}}| - \frac{1}{2}\log|HRH^T| \tag{4.25}$$

The first term in Equation (4.25) is $I(X^{k-1}, Y^{k-1})$. The second term is

$$-\frac{1}{2}\log|HRH^T| + \frac{1}{2}\log|\Sigma_{y_k|Y^{k-1}}| = -h(y_k|x_k) + h(y_k|Y^{k-1}) \tag{4.26}$$

Equation (4.25) together with Equation (4.26) shows that Equation (4.16) agrees with the derivations about the mutual information between the state and the measurement sequences for the LTI discrete-time Gaussian stochastic systems.

An interesting observation about the change of the value of the mutual information with respect to time is stated by the fact given below.

Fact 4.1.1.1:

$$I(X^k, Y^{k-1}) = I(X^{k-1}, Y^{k-1}) \tag{4.27}$$

Proof:

$$I(X^k, Y^{k-1}) = \iint f(X^k, Y^{k-1}) \log \frac{f(X^k, Y^{k-1})}{f(X^k) f(Y^{k-1})} dX^k dY^{k-1} \tag{4.28}$$

$$I(X^k, Y^{k-1}) = \iint f(x_k|x_{k-1}) f(X^{k-1}, Y^{k-1}) \log \frac{f(x_k|x_{k-1}) f(X^{k-1}, Y^{k-1})}{f(x_k|x_{k-1}) f(X^{k-1}) f(Y^{k-1})} dX^k dY^{k-1} \tag{4.29}$$

$$I(X^k, Y^{k-1}) = \iint f(x_k|x_{k-1}) f(X^{k-1}, Y^{k-1}) \log \frac{f(X^{k-1}, Y^{k-1})}{f(X^{k-1}) f(Y^{k-1})} dX^k dY^{k-1} \tag{4.30}$$

$$I(X^k, Y^{k-1}) = \iint f(X^{k-1}, Y^{k-1}) \log \frac{f(X^{k-1}, Y^{k-1})}{f(X^{k-1}) f(Y^{k-1})} dX^{k-1} dY^{k-1} \tag{4.31}$$

which is equal to $I(X^{k-1}, Y^{k-1})$. ■

The fact proved above indicates that the information gained about the state sequence does not change without the new information, i.e., a new measurement. Following derivations confirms this fact for LTI discrete-time Gaussian stochastic systems.

By using Equation (4.17) for the linear time invariant systems, the mutual information between the state and the measurement sequences can be written as:

$$I(X^k, Y^{k-1}) = \frac{1}{2} \log \frac{|\Sigma_{X^k}| |\Sigma_{Y^{k-1}}|}{|\Sigma_{[X^k, Y^{k-1}]}|} \tag{4.32}$$

The above equation and

$$|\Sigma_{[X^k, Y^{k-1}]}| = |\Sigma_{[X^{k-1}, Y^{k-1}]}| \, |GQG^T| \tag{4.33}$$

lead to the following result

$$I(X^k, Y^{k-1}) = \frac{1}{2} \log \frac{|\Sigma_{X^{k-1}}| |GQG^T| |\Sigma_{Y^{k-1}}|}{|\Sigma_{[X^{k-1}, Y^{k-1}]}| |GQG^T|} = \frac{1}{2} \log \frac{|\Sigma_{X^{k-1}}| |\Sigma_{Y^{k-1}}|}{|\Sigma_{[X^{k-1}, Y^{k-1}]}|} \tag{4.34}$$

which is equal to $I(X^{k-1}, Y^{k-1})$. This result shows that Equation (4.27) agrees with the previous derivations about mutual information between the state and the measurement sequences for the LTI discrete-time Gaussian stochastic systems.

4.1.2 Recursive Evaluation of the Observability Measure Based on the Last State

In this subsection, the observability measure based on the mutual information between the last state and the measurement sequence $I(x_k, Y^k)$ is expressed in a recursive manner.

Theorem 4.2: The observability measure based on the mutual information between the last state and the measurement sequence $I(x_k, Y^k)$ can be expressed in a recursive manner as:

$$\begin{aligned} I(x_k, Y^k) = I(x_{k-1}, Y^{k-1}) + h(x_k|x_{k-1}) - h(x_{k-1}|x_k) + h(x_{k-1}|Y^{k-1}) \\ - h(x_k|Y^{k-1}) - h(y_k|x_k) + h(y_k|Y^{k-1}) \end{aligned} \tag{4.35}$$

Proof: The mutual information between the last state and the measurement sequence $I(x_k, Y^k)$ can be expressed by using the relationships between the entropy and the mutual information (2.15),

$$I(x_k, Y^k) = h(x_k) + h(Y^k) - h(x_k, Y^k) \tag{4.36}$$

$$I(x_k, Y^k) = h(x_k) - h(x_k|Y^k) \tag{4.37}$$

$$I(x_{k-1}, Y^{k-1}) = h(x_{k-1}) + h(Y^{k-1}) - h(x_{k-1}, Y^{k-1}) \tag{4.38}$$

By adding and subtracting $I(x_{k-1}, Y^{k-1})$ to and from Equation (4.36), one can get

$$\begin{aligned} I(x_k, Y^k) = h(x_k) + h(Y^k) - h(x_k, Y^k) + I(x_{k-1}, Y^{k-1}) - h(x_{k-1}) - h(Y^{k-1}) \\ + h(x_{k-1}, Y^{k-1}) \end{aligned} \tag{4.39}$$

By using Equation (2.3), $h(Y^k)$ and $h(x_k)$ can be written as

$$h(Y^k) = h(Y^{k-1}) + h(y_k|Y^{k-1}) \tag{4.40}$$

$$h(x_k) = h(x_{k-1}) + h(x_k|x_{k-1}) - h(x_{k-1}|x_k) \tag{4.41}$$

Substituting Equations (4.40) and (4.41) into Equation (4.39) and arranging the terms, one can get the following relation:

$$\begin{aligned} I(x_k, Y^k) = I(x_{k-1}, Y^{k-1}) + h(x_k|x_{k-1}) - h(x_{k-1}|x_k) + h(x_{k-1}|Y^{k-1}) \\ - h(x_k|Y^{k-1}) - h(y_k|x_k) + h(y_k|Y^{k-1}) \end{aligned} \tag{4.42}$$

■

A similar expression for the mutual information is also given in (Bansal and Basar (1989)). Their result is given below as Fact 4.1.2.1. We prove this fact once again to show the relationship between the observability measures of the last sate and the state sequence using our framework.

Fact 4.1.2.1: (Bansal and Basar (1989))

$$I(x_k, Y^k) = I(x_k, Y^{k-1}) - h(y_k|x_k) + h(y_k|Y^{k-1}) \tag{4.43}$$

Proof: By using Equation (2.15), one can write the following relation:

$$I(x_k, Y^{k-1}) = h(x_k) + h(Y^{k-1}) - h(x_k, Y^{k-1}) \tag{4.44}$$

By adding and subtracting $I(x_k, Y^{k-1})$ to and from Equation (4.36), one can get

$$I(x_k, Y^k) = h(x_k) + h(Y^k) - h(x_k, Y^k) + I(x_k, Y^{k-1}) - h(x_k) - h(Y^{k-1}) + h(x_k, Y^{k-1}) \tag{4.45}$$

$$I(x_k, Y^k) = I(x_k, Y^{k-1}) + h(Y^k) - h(Y^{k-1}) + h(x_k, Y^{k-1}) - h(x_k, Y^k) \tag{4.46}$$

By using Equation (2.3), one can write the following relations:

$$h(Y^k) - h(Y^{k-1}) = h(y_k | Y^{k-1}) \tag{4.47}$$

$$h(x_k, Y^{k-1}) - h(x_k, Y^k) = -h(y_k | x_k) \tag{4.48}$$

By substituting Equations (4.47) and (4.48) into Equation (4.46), one can get the following result:

$$I(x_k, Y^k) = I(x_k, Y^{k-1}) - h(y_k | x_k) + h(y_k | Y^{k-1}) \tag{4.49}$$

■

Note that the increment from $I(x_k, Y^{k-1})$ to $I(x_k, Y^k)$ is exactly equal to the increment from $I(X^{k-1}, Y^{k-1})$ to $I(X^k, Y^k)$. This comes from the following fact.

Fact 4.1.2.2:

$$I(X^k, y_k) = I(x_k, y_k) \tag{4.50}$$

Proof:

$$I(X^k, y_k) = \iint f(X^k, y_k) \log \frac{f(X^k, y_k)}{f(X^k) f(y_k)} dX^k dy_k \tag{4.51}$$

$$I(X^k, y_k) = \iint f(X^k, y_k) \log \frac{f(y_k | x_k) f(X^k)}{f(X^k) f(y_k)} dX^k dy_k \tag{4.52}$$

$$I(X^k, y_k) = \iint f(x_k, y_k) \log \frac{f(y_k | x_k)}{f(y_k)} dx_k dy_k \tag{4.53}$$

$$I(X^k, y_k) = \iint f(x_k, y_k) \log \frac{f(x_k, y_k)}{f(x_k) f(y_k)} dx_k dy_k = I(x_k, y_k) \tag{4.54}$$

■

The fact tells that the information increase due to the new measurement y_k is directly related with x_k, and the possible increase in the information of the previous states is implicit and related to the information of the last state.

Fact 4.1.2.3:

$$\mathrm{I}\left(x_k, Y^{k-1}\right) = \mathrm{I}\left(x_{k-1}, Y^{k-1}\right) + \mathrm{h}(x_k|x_{k-1}) - \mathrm{h}(x_{k-1}|x_k) + \mathrm{h}\left(x_{k-1}\middle|Y^{k-1}\right) - \mathrm{h}\left(x_k\middle|Y^{k-1}\right) \tag{4.55}$$

Proof: By using Equations (4.43) and (4.35), the proof is trivial. ■

Equivalent recursive expressions are obtained below for the linear time invariant systems. To verify the relations given in Equations (4.35), (4.43) and (4.55) we refer to Chapter 3. The mutual information expressions derived in Chapter 3 are as follows:

$$\mathrm{I}\left(x_k, Y^k\right) = \frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|}{\left|\Sigma_{x_k|Y^k}\right|} \tag{4.56}$$

$$\mathrm{I}\left(x_k, Y^k\right) = \frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|}{\left|\Sigma_{x_k|Y^{k-1}}\right|} + \frac{1}{2}\log\frac{\left|\Sigma_{Y^k}\right|}{\left|\Sigma_{Y^{k-1}}\right|\left|HRH^T\right|} \tag{4.57}$$

In Equation (4.57), the first term is $\mathrm{I}\left(x_k, Y^{k-1}\right)$. The second term is

$$\mathrm{h}\left(y_k\middle|Y^{k-1}\right) - \mathrm{h}(y_k|x_k) = \mathrm{h}\left(Y^k\right) - \mathrm{h}\left(Y^{k-1}\right) - \mathrm{h}(y_k|x_k) = \frac{1}{2}\log\frac{\left|\Sigma_{Y^k}\right|}{\left|\Sigma_{Y^{k-1}}\right|\left|HRH^T\right|} \tag{4.58}$$

By using Equation (4.56), one can write the following relation:

$$\mathrm{I}\left(x_k, Y^{k-1}\right) = \frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|}{\left|\Sigma_{x_k|Y^{k-1}}\right|} = \frac{1}{2}\log\frac{\left|\Sigma_{x_{k-1}}\right|}{\left|\Sigma_{x_{k-1}|Y^{k-1}}\right|} + \frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|\left|\Sigma_{x_{k-1}|Y^{k-1}}\right|}{\left|\Sigma_{x_k|Y^{k-1}}\right|\left|\Sigma_{x_{k-1}}\right|} \tag{4.59}$$

The first term in Equation (4.59) is $\mathrm{I}\left(x_{k-1}, Y^{k-1}\right)$. By using Equation (2.3), one can write the following relation:

$$\begin{aligned}\mathrm{h}(x_k|x_{k-1}) - \mathrm{h}(x_{k-1}|x_k) + \mathrm{h}\left(x_{k-1}\middle|Y^{k-1}\right) - \mathrm{h}\left(x_k\middle|Y^{k-1}\right) \\ = \mathrm{h}(x_k) - \mathrm{h}(x_{k-1}) + \mathrm{h}\left(x_{k-1}\middle|Y^{k-1}\right) - \mathrm{h}\left(x_k\middle|Y^{k-1}\right)\end{aligned} \tag{4.60}$$

which is equal to the second term in Equation (4.59) for the LTI discrete-time Gaussian stochastic systems.

Hence Equations (4.35), (4.43) and (4.55) agree with the derivations given for the LTI discrete-time Gaussian stochastic systems.

Below we give a very short summary of the basics of particle filtering. This part is by no means a complete description of the filter but is put here for the sake of completeness.

4.2 Particle Filter

Filtering is the recursive computation of the conditional pdf of the state. Bayes theory gives the necessary equations that are used for this purpose. Here we will briefly write the recursive expressions for the pdf of the state conditioned on the given measurements and explain again very briefly how the particle filter approach solves the computation problem. Ristic et al. (2004) and Arulampalam et al. (2002) are good references for the particle filtering.

In this part, for simplicity, it is assumed that the process and measurement noises are additive so the nonlinear stochastic system is represented by:

$$x_{k+1} = p(x_k) + w_k \tag{4.61}$$

$$y_k = g(x_k) + v_k \tag{4.62}$$

Note that the aim in filtering is to find $f(x_k|Y^k)$ recursively. The following is the derivation of the necessary equations.

In Bayes filtering, the following equations are given for filtering (Equation (4.63)) and prediction (Equation (4.64)).

$$f(x_k|Y^k) = \frac{f(y_k|x_k)f(x_k|Y^{k-1})}{f(y_k|Y^{k-1})} \propto f(y_k|x_k)f(x_k|Y^{k-1}) \tag{4.63}$$

$$f(x_k|Y^{k-1}) = \int f(x_k|x_{k-1})\, f(x_{k-1}|Y^{k-1})dx_{k-1} \tag{4.64}$$

Analytical evaluation of the above two equations is possible only for very limited cases. One well known case is the Linear Gaussian system that can be solved analytically by Kalman filter. When the system is nonlinear it is almost impossible to find an analytic solution. The particle filtering is a numerical method that can give a numerical solution to the problem. Here it will be very briefly explained the most primitive but much used form of the filter namely SIR (Sequential Importance Resampling) algorithm.

SIR algorithm is based on the following approximation:

$$f(x_{k-1}|Y^{k-1}) \approx \sum_{i=1}^{M} s_{k-1}^{(i)} \delta(x_{k-1} - x_{k-1}^{(i)}) \tag{4.65}$$

In the above expression $x_{k-1}^{(i)}$ is the last element of a state sequence which is called a 'particle'. δ is the delta Dirac function. Prediction step consists of computation of $f(x_k|Y^{k-1})$. This can be achieved by using the prediction equation and the approximate $f(x_{k-1}|Y^{k-1})$ given in Equation (4.65) as follows:

$$f\left(x_k|Y^{k-1}\right) = \int f(x_k|x_{k-1}) \sum_{i=1}^{M} s_{k-1}^{(i)} \delta(x_{k-1} - x_{k-1}^{(i)})\, dx_{k-1} = \sum_{i=1}^{M} s_{k-1}^{(i)} f\left(x_k \middle| x_{k-1}^{(i)}\right) \qquad (4.66)$$

To continue with the filtering part we need to represent $f\left(x_k \middle| Y^{k-1}\right)$ approximately by impulses again. The requirement is satisfied if we draw samples from the prediction density. Another way of drawing samples from the prediction density is to draw one sample from the process noise and add it to $p(x_{k-1}^{(i)})$ for each (i), i.e., $x_k^{(i)} = p(x_{k-1}^{(i)}) + w_{k-1}^{(i)}$. Prediction density is obtained as the following sum.

$$f\left(x_k|Y^{k-1}\right) = \sum_{i=1}^{M} s_{k-1}^{(i)} \delta(x_k - x_k^{(i)}) \qquad (4.67)$$

The above method is corresponds to the selection of the importance density as the prior. So that the new samples are drawn from the prior, i.e., $f\left(x_k|Y^{k-1}\right)$. Using the prior as the importance density is not the optimal use of the available measurements. However, since the aim is to only demonstrate the use of the particle filter in the computation of the observability measure the subject will not elaborated from this point of view.

The filtering step of the algorithm can be considered as recomputation of the weights of the particles according to the fit of the measurement to the predicted density. Using the filtering equation of the Bayes filter the following relations are obtained:

$$f\left(x_k^{(i)} \middle| Y^k\right) \propto f\left(y_k|x_k^{(i)}\right) s_{k-1}^{(i)} \rightarrow \tilde{s}_k^{(i)} = f\left(y_k|x_k^{(i)}\right) s_{k-1}^{(i)} \qquad (4.68)$$

The new weights obtained above should be normalized to have the sum equal to one. Normalized weights are denoted by $s_k^{(i)}$.

The above algorithm is named as SIS (Sequential Importance Sampling) algorithm. The algorithm is summarized in Table 4.1 below.

Table 4.1 Filtering via SIS

$\left[\left\{x_k^{(i)}, s_k^{(i)}\right\}_{i=1}^{N}\right] = \text{SIS}\left[\left\{x_{k-1}^{(i)}, s_{k-1}^{(i)}\right\}_{i=1}^{N}, y_k\right]$

- FOR $i = 1{:}N$
    - Draw $x_k^{(i)} \sim f\left(x_k^{(i)}|x_{k-1}^{(i)}\right)$
    - Evaluate the importance weights $\tilde{s}_k^{(i)} = s_{k-1}^{(i)} f\left(y_k|x_k^{(i)}\right)$
- END FOR

- Calculate total weight: $\mathrm{t} = \mathrm{SUM}\left[\left\{\tilde{\mathrm{s}}_{\mathrm{k}}^{(\mathrm{i})}\right\}_{\mathrm{i}=1}^{\mathrm{N}}\right]$
- FOR $\mathrm{i} = 1{:}\,\mathrm{N}$
    - Normalize: $\mathrm{s}_{\mathrm{k}}^{(\mathrm{i})} = \mathrm{t}^{-1}\tilde{\mathrm{s}}_{\mathrm{k}}^{(\mathrm{i})}$
- END FOR

The main drawback of the algorithm is its divergence problem: after few time steps the weights of almost all of the particles reduce to very small numbers except for few. The resultant sum of the impulses becomes insufficient to represent the underlying density. To overcome this problem resampling is done and 'good' particles are multiplied and 'bad' particles are eliminated either at each time or whenever necessary. There are several methods to resample the particles. One well known method which has low computation cost is Systematic Resampling. This algorithm first obtains the cumulative distribution of the weights. A real number a is drawn from the uniform distribution $\mathrm{U}(0, 1/\mathrm{N})$. At each time this number is increased by an amount $1/\mathrm{N}$ and the value obtained is compared with the cumulative distribution of the weights. The overall algorithm is given in Table 4.2.

Table 4.2 Resampling Algorithm

$\left[\left\{\mathrm{x}_{\mathrm{k}}^{(\mathrm{j})^*}, \mathrm{s}_{\mathrm{k}}^{(\mathrm{j})}, \mathrm{i}^{(\mathrm{j})}\right\}_{\mathrm{j}=1}^{\mathrm{N}}\right] = \mathrm{RESAMPLE}\left[\left\{\mathrm{x}_{\mathrm{k}}^{(\mathrm{i})}, \mathrm{s}_{\mathrm{k}}^{(\mathrm{i})}\right\}_{\mathrm{i}=1}^{\mathrm{N}}\right]$

- Initialize the CSW: $\mathrm{c}_1 = \mathrm{s}_{\mathrm{k}}^{1}$
- FOR $\mathrm{i} = 2{:}\,\mathrm{N}$
    - Construct CSW: $\mathrm{c}_{\mathrm{i}} = \mathrm{c}_{\mathrm{i}-1} + \mathrm{s}_{\mathrm{k}}^{(\mathrm{i})}$
- END FOR
- Start at the bottom of the CSW: $\mathrm{i} = 1$
- Draw a starting point: $\mathrm{u}_1 \sim \mathrm{U}(0, 1/\mathrm{N})$
- FOR j= $1{:}\,\mathrm{N}$
    - Move along the CSW: $\mathrm{u}_{\mathrm{j}} = \mathrm{u}_1 + (\mathrm{j} - 1)/\mathrm{N}$
    - WHILE $\mathrm{u}_{\mathrm{j}} > \mathrm{c}_{\mathrm{i}}$
        - $\mathrm{i} = \mathrm{i} + 1$
    - END WHILE
    - Assign sample: $\mathrm{x}_{\mathrm{k}}^{(\mathrm{j})^*} = \mathrm{x}_{\mathrm{k}}^{(\mathrm{i})}$
    - Assign weight: $\mathrm{s}_{\mathrm{k}}^{(\mathrm{j})} = 1/\mathrm{N}$
    - Assign parent: $\mathrm{i}^{(\mathrm{j})} = \mathrm{i}$

- END FOR

4.3 Calculation of the Observability Measures by Using the Particle Filters

In this section, the recursive calculations of the observability measures are presented by using the particle filters. In the first subsection, the calculation of the observability measure based on the mutual information between the state and the measurement sequences is considered. The calculation of the observability measure based on the last state and the measurement sequence is given in the second subsection. At the end of the section, we provide an example by applying these calculations to a simple LTI discrete-time Gaussian stochastic system.

4.3.1 Calculation of the Observability Measure Based on the State Sequence

The observability measure based on the mutual information between the state and the measurement sequences is written recursively as in Equation (4.5)

$$I\left(X^k, Y^k\right) = I\left(X^{k-1}, Y^{k-1}\right) - h(y_k|x_k) + h\left(y_k\middle|Y^{k-1}\right) \tag{4.69}$$

where $h(.)$ is the differential entropy, X^k is the state sequence and Y^k is the measurement sequence.

The differential entropies given in the above equation can be written as

$$h(y_k|x_k) = -\iint f(x_k, y_k) \log f(y_k|x_k) dx_k dy_k \tag{4.70}$$

$$h\left(y_k\middle|Y^{k-1}\right) = -\int f\left(Y^k\right) \log f\left(y_k\middle|Y^{k-1}\right) dY^k \tag{4.71}$$

$$h\left(y_k\middle|Y^{k-1}\right) = -\int f\left(Y^{k-1}\right) \int f\left(y_k\middle|Y^{k-1}\right) \log f\left(y_k\middle|Y^{k-1}\right) dy_k dY^{k-1} \tag{4.72}$$

To evaluate these differential entropies, we need the probability density functions $f(x_k, y_k)$, $f(y_k|x_k)$, $f\left(Y^k\right)$ and $f\left(y_k\middle|Y^{k-1}\right)$. The state conditional probability density function of the output, $f(y_k|x_k)$ is a known function that determines the measurement equation. The joint pdf of the last state and the last output, $f(x_k, y_k)$, can be calculated as

$$f(x_k, y_k) = f(y_k|x_k) f(x_k) \tag{4.73}$$

In Equation (4.73) $f(x_k)$ can be calculated recursively as

$$f(x_k) = \int f(x_k|x_{k-1})f(x_{k-1})dx_{k-1} \tag{4.74}$$

where $f(x_k|x_{k-1})$ is the conditional pdf that represents the dynamic structure of the system.

In Equation (4.72), the function $f(y_k|Y^{k-1})$ can be calculated from

$$f(y_k|Y^{k-1}) = \int f(y_k|x_k)f(x_k|Y^{k-1})dx_k \tag{4.75}$$

Notice that $f(x_k|Y^{k-1})$ is the prediction density in the filtering problems.

The problem that is analyzed in this thesis is not the computation of the estimated states but the mutual information between the states and the measurements. This requires the computation of the unconditional pdf of the state. The computation is done by Monte Carlo methods. The required pdf is represented by

$$f(x_{k-1}) \approx \sum_{i=1}^{N} \frac{1}{N} \delta\left(x_{k-1} - x_{k-1}^{(i)}\right) \tag{4.76}$$

and the recursion equation is obtained by using the Bayes theory as given in Equation (4.74) as:

$$f(x_k) = \int f(x_k|x_{k-1})f(x_{k-1})dx_{k-1} = \sum_{i=1}^{N} \frac{1}{N} f\left(x_k|x_{k-1}^{(i)}\right) \tag{4.77}$$

To complete the cycle new N samples are drawn from this distribution and the distribution is approximated as before,

$$f(x_k) \approx \sum_{i=1}^{N} \frac{1}{N} \delta\left(x_k - x_k^{(i)}\right) \tag{4.78}$$

The conditional densities of the state are obtained by the particle filtering. These densities are assumed to have the representations given below.

$$f(x_k|Y^{k-1}) \approx \sum_{i=1}^{N} s_{k|k-1}^{(i)} \delta\left(x_k - x_{k|k-1}^{(i)}\right) \tag{4.79}$$

$$f(x_k|Y^{k}) \approx \sum_{i=1}^{N} s_{k|k}^{(i)} \delta\left(x_k - x_{k|k}^{(i)}\right) \tag{4.80}$$

The particle filter gives the relationship between $\left\{s_{k|k-1}^{(i)}, x_{k|k-1}^{(i)}\right\}$ and $\left\{s_{k|k}^{(i)}, x_{k|k}^{(i)}\right\}$.

To find the incremental change in the observability measure it is necessary to calculate $-h(y_k|x_k) + h(y_k|Y^{k-1})$. Each term of the expression is computed by first computing $f(x_k, y_k)$. By using Equation (4.73) $f(x_k, y_k)$ can be calculated as

$$f(x_k, y_k) = f(y_k|x_k)f(x_k) \approx \sum_{i=1}^{N} \frac{1}{N} f\left(y_k \left|x_k^{(i)}\right.\right) \delta\left(x_k - x_k^{(i)}\right) \tag{4.81}$$

By substituting Equation (4.81) into Equation (4.70) the following relation can be found

$$h(y_k|x_k) \approx -\iint \left[\sum_{i=1}^{N} \frac{1}{N} f\left(y_k \left| x_k^{(i)}\right.\right) \delta\left(x_k - x_k^{(i)}\right)\right] \log f(y_k|x_k) dx_k dy_k \tag{4.82}$$

$$h(y_k|x_k) \approx -\sum_{i=1}^{N} \frac{1}{N} \int f\left(y_k \left| x_k^{(i)}\right.\right) \log f\left(y_k \left| x_k^{(i)}\right.\right) dy_k \tag{4.83}$$

The computation of $h(y_k|Y^{k-1})$ is problematic since we need $f(y_k|Y^{k-1})$ which depends on the past measurements as the given information. Here we make an approximation and replace the given information Y^{k-1} by the conditional pdf of the state. With this approximation for the inner integral one can write:

$$\begin{gathered} \int f(y_k|Y^{k-1}) \log f(y_k|Y^{k-1}) dy_k \approx \\ \int f\left(y_k \left| \left\{s_{k|k-1}^{(i)}, x_{k|k-1}^{(i)}\right\}\right.\right) \log f\left(y_k \left| \left\{s_{k|k-1}^{(i)}, x_{k|k-1}^{(i)}\right\}\right.\right) dy_k \end{gathered} \tag{4.84}$$

which becomes independent of Y^{k-1}. The outer integral then reduces to 1. Note that

$$f(y_k|Y^{k-1}) \approx \sum_{i=1}^{N} s_{k|k-1}^{(i)} f\left(y_k \left| x_{k|k-1}^{(i)}\right.\right) \tag{4.85}$$

which is not in the form of the sum of impulses but is a continuous function of y_k. Using Equations (4.84) and (4.85), Equation (4.71) can be written as

$$h(y_k|Y^{k-1}) \approx -\int \left[\sum_{i=1}^{N} s_{k|k-1}^{i} f\left(y_k \left| x_{k|k-1}^{(i)}\right.\right)\right] \log \left[\sum_{i=1}^{N} s_{k|k-1}^{i} f\left(y_k \left| x_{k|k-1}^{(i)}\right.\right)\right] dy_k \tag{4.86}$$

The computation of the update in the observability measure for the next observation is the difference of the expressions given in Equations (4.86) and (4.83). Unfortunately the computation requires integration for both of the terms. These integrals can be calculated by using numerical integration techniques (Hoffmann and Tomlin (2010), Ryan and Hedrick (2010)). Numerical integration may be again performed by the Monte Carlo integration.

4.3.2 Calculation of the Observability Measure Based on the Last State

In this section we formulate the observability measure based on the mutual information between the last state and the measurement sequence for the nonlinear systems. In this computation instead of using the recursive expressions we use the direct relationship between the mutual information and the entropy.

The mutual information between the last state and the measurements is given in Equation (4.37) as:

$$I(x_k, Y^k) = h(x_k) - h(x_k|Y^k) \tag{4.87}$$

where

$$h(x_k) = -\int f(x_k) \log f(x_k) dx_k \tag{4.88}$$

$$h\left(x_k \middle| Y^k\right) = -\iint f\left(x_k, Y^k\right) \log f\left(x_k \middle| Y^k\right) dx_k dY^k \tag{4.89}$$

In the following derivations it is assumed that the particle filter is applied to estimate the posterior density of the state. Both the prior and posterior pdf's are approximated by sum of weighted impulses.

From Equations (4.77) and (4.88) $h(x_k)$ can be expressed as:

$$\begin{aligned} h(x_k) = -\int f(x_k) \log f(x_k) dx_k = \\ -\int \left[\sum_{j=1}^{N} s_{k-1}^{(i)} f\left(x_k \middle| x_{k-1}^{(i)}\right)\right] \log \left[\sum_{i=1}^{N} s_{k-1}^{(i)} f\left(x_k \middle| x_{k-1}^{(i)}\right)\right] dx_k \end{aligned} \tag{4.90}$$

$f\left(x_k \middle| Y^k\right)$ can be calculated from the following relation:

$$f\left(x_k \middle| Y^k\right) = \frac{f(y_k|x_k)}{f(y_k|Y^{k-1})} f\left(x_k \middle| Y^{k-1}\right) \tag{4.91}$$

$h\left(x_k \middle| Y^k\right)$ can be written as

$$\begin{aligned} h\left(x_k \middle| Y^k\right) &= -\iint f\left(x_k, Y^k\right) \log f\left(x_k \middle| Y^k\right) dx_k dY^k \\ &= -\iint f\left(x_k, Y^k\right) \log \left[\frac{f(y_k|x_k)}{f(y_k|Y^{k-1})} f\left(x_k \middle| Y^{k-1}\right)\right] dx_k dY^k \end{aligned} \tag{4.92}$$

$$h\left(x_k \middle| Y^k\right) = h(y_k|x_k) - h\left(y_k \middle| Y^{k-1}\right) + h\left(x_k \middle| Y^{k-1}\right) \tag{4.93}$$

$h(y_k|x_k)$ and $h\left(y_k \middle| Y^{k-1}\right)$ are given in Equations (4.83) and (4.86), respectively.

$h\left(x_k \middle| Y^{k-1}\right)$ can be calculated from the density $f\left(x_k \middle| Y^{k-1}\right)$ as:

$$f\left(x_k \middle| Y^{k-1}\right) = \int f(x_k|x_{k-1})\, f\left(x_{k-1} \middle| Y^{k-1}\right) dx_{k-1} \tag{4.94}$$

where $f\left(x_{k-1} \middle| Y^{k-1}\right)$ is the filtering density at time $(k-1)$. Then

$$f\left(x_{k-1} \middle| Y^{k-1}\right) \approx \sum_{i=1}^{N} s_{k-1|k-1}^{(i)} \delta\left(x_{k-1} - x_{k-1|k-1}^{(i)}\right) \tag{4.95}$$

$$f\left(x_k \middle| Y^{k-1}\right) \approx \sum_{i=1}^{N} s_{k-1|k-1}^{(i)} f\left(x_k \middle| x_{k-1|k-1}^{(i)}\right) \tag{4.96}$$

$$h\left(x_k \middle| Y^{k-1}\right) = -\int f\left(Y^{k-1}\right) \int f\left(x_k \middle| Y^{k-1}\right) \log f\left(x_k \middle| Y^{k-1}\right) dx_k dY^{k-1} \tag{4.97}$$

The given information of the conditional pdf's is Y^{k-1} which can be approximated as the conditional pdf of the state, $\left\{s_{k-1|k-1}^{(i)}, x_{k-1|k-1}^{(i)}\right\}$. Using this approximation the inner integration becomes independent of Y^{k-1} and can be considered as a term which is fixed. For this case the outer integration reduces to 1. With this explanation the following expression is obtained for $h\left(x_k \middle| Y^{k-1}\right)$,

$$h\left(x_k \middle| Y^{k-1}\right) \approx -\int \left[\sum_{i=1}^{N} s_{k-1|k-1}^{(i)} f\left(x_k \middle| x_{k-1|k-1}^{(i)}\right)\right] \log \left[\sum_{i=1}^{N} s_{k-1|k-1}^{(i)} f\left(x_k \middle| x_{k-1|k-1}^{(i)}\right)\right] dx_k \tag{4.98}$$

Again Equations (4.98) and (4.90) can be calculated using numerical integration techniques.

Here the observability measures are obtained, both for the state sequence and the last state cases. Below the theory is applied to a simple linear system. We demonstrate that the particle filter approach and the theory developed for the linear systems agree with the observability measures.

Example 4.1: Let us assume that we have a LTI discrete-time Gaussian stochastic system represented by Equations given below.

$$x_{k+1} = 0.9x_k + w_k \tag{4.99}$$

$$y_k = x_k + v_k \tag{4.100}$$

where

$$x_0 \sim N(0, 1),\ w_k \sim N(0,1),\ v_k \sim N(0,1) \tag{4.101}$$

The aim of the example is to compute $I\left(X^k, Y^k\right)$ and $I\left(x_k, Y^k\right)$ both using the theory developed in Chapter 3 and the particle filter method developed in this chapter. Simulation time is taken as $k = 40$ and the particle filter is realized by 1000 particles.

In Figure 4.1, the system state obtained by the Kalman filter is compared with the one obtained by the particle filter. Since the system is linear and the noises are Gaussian the optimal estimate is the Kalman filter estimate. The similarity of the outputs of the two filters shows that, for this simple example, the particle filter with 1000 particles is sufficiently successful.

Figure 4.2 shows the increase in the value of $I\left(X^k, Y^k\right)$ as time increases. The curves also show the close relationship between the theoretical (using the theory of Chapter 3) and the computational (the particle filter) observability measures. In Figure 4.3 the value of $I\left(x_k, Y^k\right)$ (the result of Monte Carlo simulation (100 runs)), again computed in two ways are shown.

Figures 4.1-4.3 demonstrate that the observability measures obtained by using the particle filters are close to the true results obtained theoretically. $I(X^k, Y^k)$ increases with time as expected. On the other hand $I(x_k, Y^k)$ converges to a steady state value as time increases since the system is observable and stable. This result is similar to the convergence of the Kalman filter covariance matrices to finite and initial condition independent values at the steady state. Note that if the single state example were unobservable, meaning that nothing is measured about the state, obviously $I(X^k, Y^k)$ or $I(x_k, Y^k)$ would be zero. On the other hand if the system were unstable the measurements would be very valuable that makes the observability measure $I(x_k, Y^k)$ increasing to infinity. Note that if the quality of the estimation increases the calculations of the observability measures become more accurate. This accuracy can be achieved by using more particles in the particle filter.

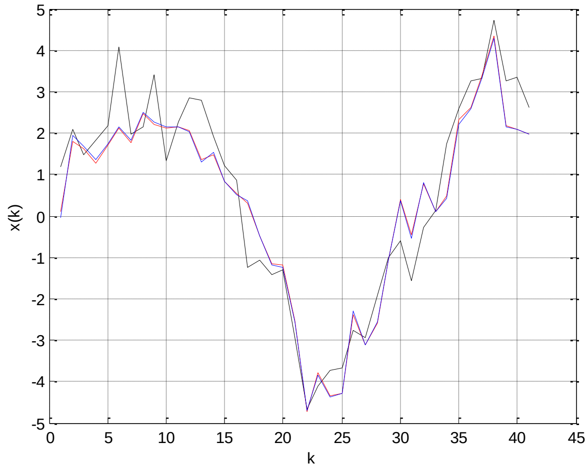

Figure 4.1 System State (True (Black), KF Estimate (Red), PF Estimate (Blue))

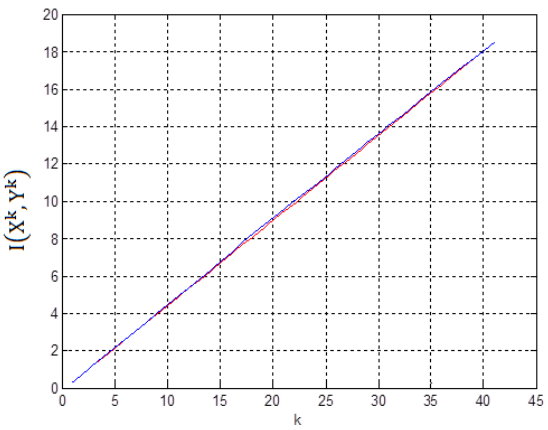

Figure 4.2 $I(X^k, Y^k)$ (Theoretical (Red), PF Calculation (Blue))

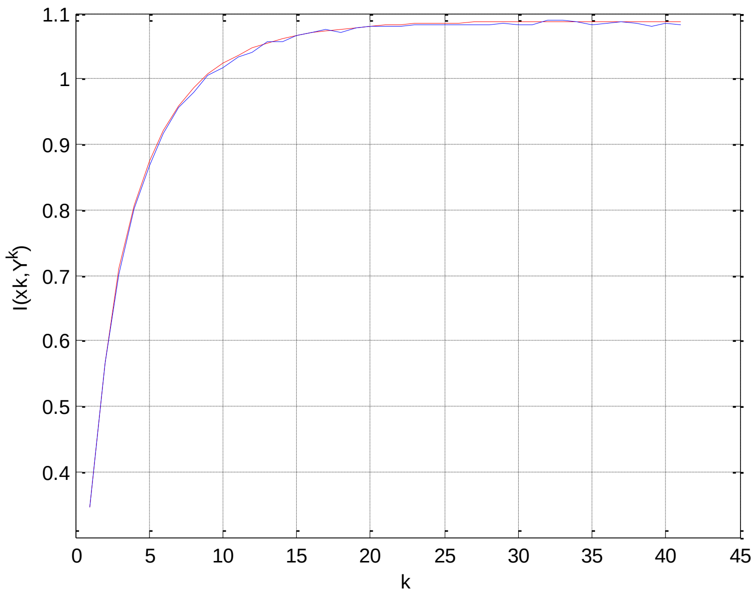

Figure 4.3 $I(x_k, Y^k)$ (Theoretical (Red), PF Calculation (Blue))

4.4 Computation of the Observability Measures to a Nonlinear Example

In this section, the two observability measures, the observability measure for the whole state sequence and the observability measure for the last state, are computed for a simple nonlinear stochastic system and the conclusions derived for LTI discrete-time Gaussian stochastic systems are investigated for nonlinear systems.

Equations of a simple one state nonlinear system are given below:

$$x_{k+1} = a_1 x_k + a_2 \frac{x_k}{1 + {x_k}^2} + w_k \tag{4.102}$$

$$y_k = c_1 x_k + c_2 {x_k}^2 + v_k \tag{4.103}$$

where $x_0 \sim N(0, \Sigma_0)$, $w_k \sim N(0, Q)$, $v_k \sim N(0, R)$, a_1, a_2, and c_2 are constants and c_1 is an increasing function of time.

We have performed six simulations using different sets of parameters. Case 1 can be considered as the baseline experiment. The parameters that are different than the baseline experiment are shown as bold for the remaining cases.

Case 1:

$$a_1 = 0.7,\ a_2 = 25,\ c_1(k) = \frac{0.99(k-1)}{40} + 0.01,\ c_2 = 0.05,\ \Sigma_0 = 5,\ Q = 10,\ R = 0.1 \qquad (4.104)$$

Case 2:

$$a_1 = 0.7,\ a_2 = 25,\ c_1(k) = \frac{0.99(k-1)}{40} + 0.01,\ \mathbf{c_2 = 0.005},\ \Sigma_0 = 5,\ Q = 10,\ R = 0.1 \qquad (4.105)$$

Case 3:

$$a_1 = 0.7,\ \mathbf{a_2 = 1},\ c_1 = \frac{0.99(k-1)}{40} + 0.01,\ c_2 = 0.05,\ \Sigma_0 = 5,\ Q = 10,\ R = 0.1 \qquad (4.106)$$

Case 4:

$$a_1 = 0.7,\ a_2 = 25,\ c_1 = \frac{0.99(k-1)}{40} + 0.01,\ c_2 = 0.05,\ \Sigma_0 = 5,\ Q = 10,\ \mathbf{R = 1} \qquad (4.107)$$

Case 5:

$$a_1 = 0.7,\ a_2 = 25,\ c_1 = \frac{0.99(k-1)}{40} + 0.01,\ c_2 = 0.05,\ \Sigma_0 = 5,\ \mathbf{Q = 1},\ R = 0.1 \qquad (4.108)$$

Case 6: s

$$a_1 = 0.7,\ a_2 = 25,\ c_1 = \frac{0.99(k-1)}{40} + 0.01,\ c_2 = 0.05,\ \boldsymbol{\Sigma_0 = 0.05},\ Q = 10,\ R = 0.1 \qquad (4.109)$$

Case 2 differs from Case 1 only in the value of c_2. c_2 is decreased in Case 2 compared to Case 1 by a factor of 10 so the effect of the nonlinear term in the measurement equation is decreased as well as the observation gain.

Case 3 differs from Case 1 only in the value of a_2. a_2 is decreased which causes a decrease in the effect of the nonlinearity in the state equation as well as the gain related with the previous value of the state.

Case 4 differs from Case 1 only in the value of the measurement noise covariance R which is increased. Case 5 differs from Case 1 only in the value of the process noise covariance Q which is decreased. Case 6 differs from Case 1 only in the value of the initial state uncertainty Σ_0 which is decreased.

Experiments are performed for all six cases; however only for the first three cases the detailed results are given. The results of the last three experiments are summarized in Table 4.4.

The Sequential Importance Resampling (SIR) filter is used in the simulations. 5000 particles are used to represent the probability density functions to avoid the poor estimation performance.

The results for Case 1 are shown in the Figures 4.4-4.7. Figure 4.4 shows the performance of the state estimation. Since c_2 increases with time the quality of the estimation increases as well. Figure 4.5 shows that $I(X^k, Y^k)$ increases with time as expected. It may be interesting to see the difference between $[I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})]$ and $I(x_k, Y^k)$. This comparison is given in Figure 4.6. We have included into this figure the result of 'the second moment approximation' method of Mohler and Hwang (1988). The second moment approximation is calculated from the following equations (Ryan and Hedrick (2010)). Equations given below simply approximates $I(x_k, Y^k)$ by the corresponding mutual information expression obtained for linear systems.

$$I(x_k, Y^k) \approx \frac{1}{2}\log\frac{|\Sigma_{x_k}|}{|\Sigma_{x_k|Y^k}|} \tag{4.110}$$

$$\Sigma_{x_k} \approx \sum_{i=1}^{N}\left(x_k^{(i)} - \mu_{x_k}\right)/(N-1) \tag{4.111}$$

$$\Sigma_{x_k|Y^k} \approx \sum_{i=1}^{N}\left(x_{k|k}^{(i)} - \mu_{x_{k|k}}\right)/(N-1) \tag{4.112}$$

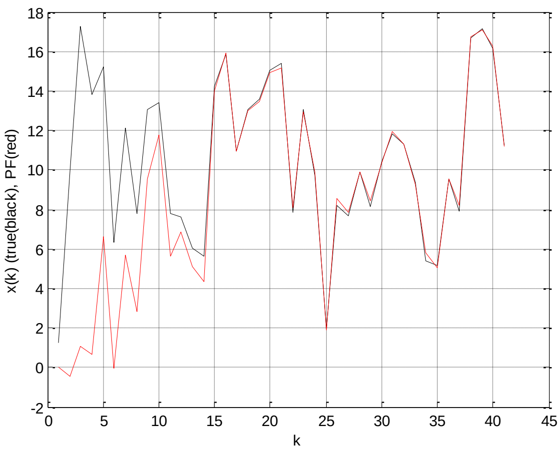

Figure 4.4 Particle Filter Estimation of the state for Case 1 (True (Black), Estimated (Red))

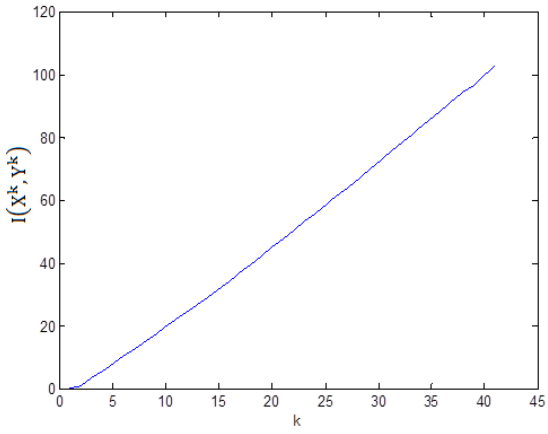

Figure 4.5 $I(X^k, Y^k)$ for Case 1

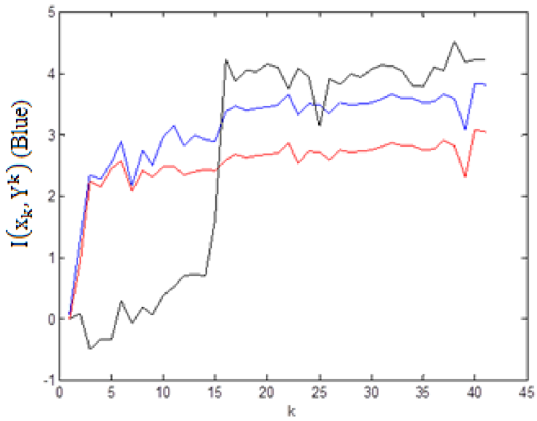

Figure 4.6 The observability measures between the last state and the measurements: The blue line is $I(x_k, Y^k)$; Red line is $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$; Black line is the Second Moment Approximation for Case 1

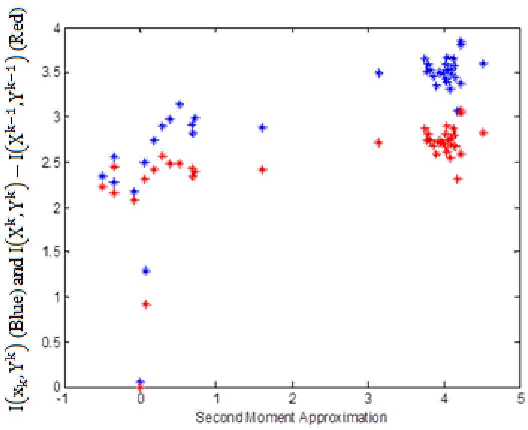

Figure 4.7 Correlation plot of the second moment approximation and the observability measures for Case 1. Blue: $I(x_k, Y^k)$ vs. the Second Moment Approximation. Red: Difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$ vs. the Second Moment Approximation

As seen from Figure 4.6 $I(x_k, Y^k)$ and the difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$ increases with time as expected; they have similar characteristics. However since the nonlinearity in the system is large, especially at the beginning, the second moment approximation deviates from $I(x_k, Y^k)$ and does not have a similar characteristics as the characteristics of $I(x_k, Y^k)$.

The results of Case 2 can be seen in the Figures 4.8-4.9. Since the nonlinearity in the measurement equation is reduced system becomes more 'linear' and the second moment approximation gives similar results as $I(x_k, Y^k)$ and/or the difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$.

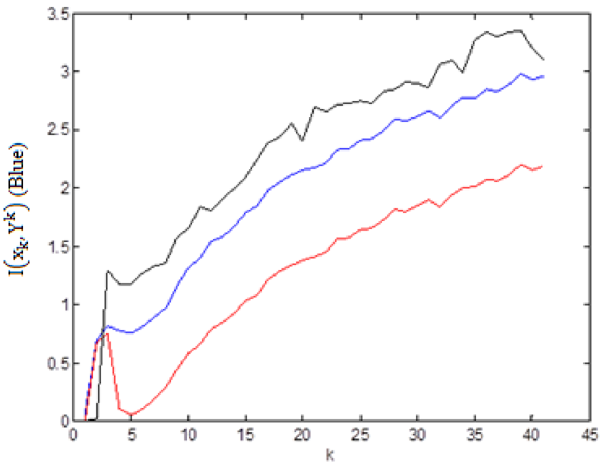

Figure 4.8 The observability measures between the last state and the measurements: The blue line is $I(x_k, Y^k)$; Red line is $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$; Black line is the Second Moment Approximation for Case 2

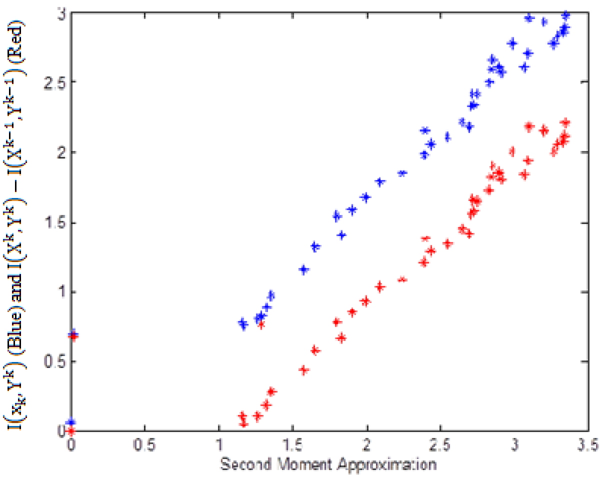

Figure 4.9 Correlation plot of the second moment approximation and the observability measures for Case 2. Blue: $I(x_k, Y^k)$ vs. the Second Moment Approximation. Red: Difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$ vs. the Second Moment Approximation

The results of Case 3 can be seen in the Figures 4.10-4.11. For this case the measurement nonlinearity is the same as Case 1 however the coefficient of the nonlinear term in the state dynamics is reduced by 1/25. Since the nonlinearity in the state equation is decreased the second moment approximation becomes closer to the $I(x_k, Y^k)$. Note that comparison of the cases 2 and 3 with Case 1 shows the effect of the nonlinearity of the measurement equation: higher measurement nonlinearity produces more diverse results for the second moment approximation.

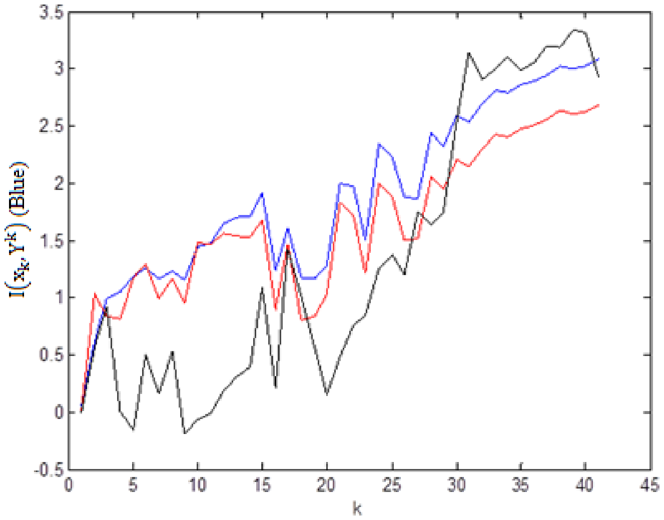

Figure 4.10 The observability measures between the last state and the measurements: The blue line is $I(x_k, Y^k)$; Red line is $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$; Black line is the Second Moment Approximation for Case 3

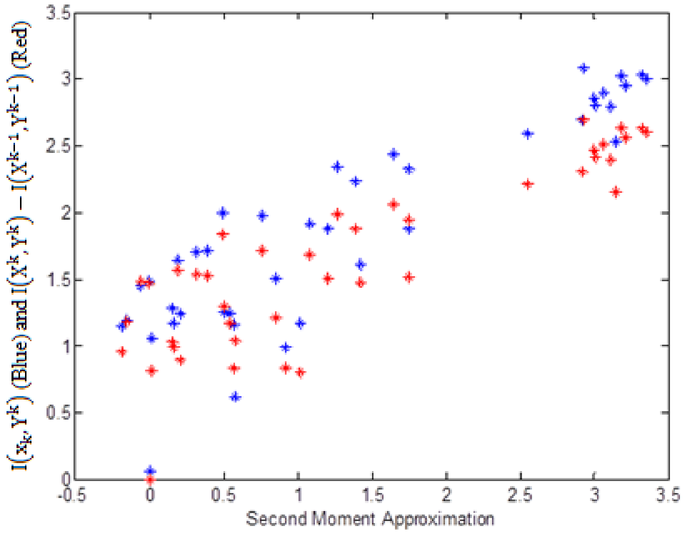

Figure 4.11 Correlation plot of the second moment approximation and the observability measures for Case 3. Blue: $I(x_k, Y^k)$ vs. the Second Moment Approximation. Red: Difference $I(X^k, Y^k) - I(X^{k-1}, Y^{k-1})$ vs. the Second Moment Approximation

The correlation coefficients between the second moment approximation and $\mathrm{I}(x_k, Y^k)$ or the difference $\mathrm{I}(X^k, Y^k) - \mathrm{I}(X^{k-1}, Y^{k-1})$ can be seen in Table 4.3. As seen from the figures and Table 4.3, when the nonlinearity increases the observability measures differs from the second moment approximation. In the most nonlinear case, which is Case 1, the correlation values are smaller than the other cases. In the least nonlinear case, which is Case 2, the figures of the observability measures versus the second moment approximation are less scattered.

The effects of system parameters given in Subasi and Demirekler (2011) for LTI discrete-time Gaussian stochastic systems are also observed for the nonlinear case. Observability measure of Case 2 is much smaller than the observability measure of Case 1. This is due to the decrease in the coefficient of the state in the measurement equation (like a noise power increase in the measurement), The system is much less 'observable'. If we compare the measures for the Case 1 and Case 3, the measures again decrease by a decrease in the coefficient of the state in the state equation. The effect of the noise terms are given in Table 4.4. in the rows 4-6. The fourth row of Table 4.4, corresponding to Case 4, shows that an increase in the measurement noise cause a decrease in the observability measures. The fifth row of the table, corresponding to Case 5, indicates that by decreasing the process noise the observability measures also decrease. The sixth row, Case 6, shows the relationship between the initial state uncertainty and the observability measures. A decrease in the initial state uncertainty causes an decrease in the measures. However, the effect of the initial state uncertainty is limited, especially for large k.

Table 4.3 Cross Correlation Coefficient Values

	Case 1	Case 2	Case 3
Cross correlation coefficients between $\mathrm{I}(x_k, Y^k)$ and the second moment approximation	0.7627	0.9756	0.8886
Cross correlation coefficients between the difference $\mathrm{I}(X^k, Y^k) - \mathrm{I}(X^{k-1}, Y^{k-1})$ and the second moment approximation	0.6007	0.9250	0.8637

Table 4.4 Effects of System Parameters

$k = 40$	$I(X^k, Y^k)$	$I(x_k, Y^k)$
$a_1 = 0.7$ $a_2 = 25$ $c_1 = \frac{0.99(k-1)}{40} + 0.01$ $c_2 = 0.05$ $\Sigma_0 = 5$ $Q = 10$ $R = 0.1$ (Case 1)	102.9626	3.8127
$a_1 = 0.7$ $a_2 = 25$ $c_1 = \frac{0.99(k-1)}{40} + 0.01$ $c_2 = 0.005$ $\Sigma_0 = 5$ $Q = 10$ $R = 0.1$ (Case 2)	52.6770	2.9593
$a_1 = 0.7$ $a_2 = 1$ $c_1 = \frac{0.99(k-1)}{40} + 0.01$ $c_2 = 0.05$ $\Sigma_0 = 5$ $Q = 10$ $R = 0.1$ (Case 3)	67.9165	3.0871
$a_1 = 0.7$ $a_2 = 25$ $c_1 = \frac{0.99(k-1)}{40} + 0.01$ $c_2 = 0.05$ $\Sigma_0 = 5$ $Q = 10$ $R = 1$ (Case 4)	62.2956	2.5881
$a_1 = 0.7$ $a_2 = 25$ $c_1 = \frac{0.99(k-1)}{40} + 0.01$ $c_2 = 0.05$ $\Sigma_0 = 5$ $Q = 1$ $R = 0.1$ (Case 5)	60.1768	2.6094
$a_1 = 0.7$ $a_2 = 25$ $c_1 = \frac{0.99(k-1)}{40} + 0.01$ $c_2 = 0.05$ $\Sigma_0 = 0.05$ $Q = 10$ $R = 0.1$ (Case 6)	99.0484	3.5944

4.5 Discussion

In this study the observability measure definitions based on the mutual information are expressed in a recursive manner. An algorithm is generated for the computation of both of the observability measures using the particle filters. The recursive expressions obtained for these measures plays an important role in the proposed algorithm. The algorithm is tested on a linear system and the observability measures obtained by the recursive algorithm are compared to the ones given theoretically as derived in Chapter 3.

The proposed algorithm is also applied to a simple nonlinear system which is a variation of a commonly used nonlinear system in the literature. The results of this example are also compared with the results of the 'second moment' approximation of Mohler and Hwang (1988).

From the experiments that we have conducted on the nonlinear system the following conclusions could be deduced:

- The effects of the changes in the parameters of the nonlinear system are very similar to the effects of the changes in the system matrices of linear systems.

- The Second moment approximation is not very suitable for highly nonlinear systems. This conclusion is due to the observed increase in the discrepancy between mutual information and its second moment approximation.

CHAPTER 5

CONCLUSIONS

In this study two observability measure definitions are given. The first measure considers the relationship between the output sequence of a stochastic system and its state sequence. The second one is more traditional and is based on the relationship between the measurement sequence and the last state. The observability measures used in this study are the mutual information, and the Bhattacharyya and the Hellinger distances.

The mutual information between the state and the measurement sequences as an observability measure of a stochastic system is proposed for the first time in this thesis. This observability measure is examined for LTI discrete-time Gaussian stochastic systems in detail and related analytical expressions are obtained. The measure is derived in terms of the statistics of the basic random variables and the system matrices. The effects of the system matrices, initial state uncertainty, and the process and the measurement noises on the observability measure are examined. Two interesting results are observed:

1. The unobservable states of the deterministic system has no effect on the measure
2. Any observable part with no measurement uncertainty makes the measure infinite

Since the mutual information is a special case of Kullback-Leibler distance, other probabilistic distance measures, the Bhattacharyya and the Hellinger distances, are also investigated in full detail to be used as observability measures. These measures show results that are similar to the ones obtained using the mutual information.

The relationship between the observability measures and the covariance matrices of the states conditioned on the measurements are derived explicitly. The mutual information compares the determinants of the conditional (to the measurements) and the unconditional covariance matrices of the state sequence.

The second observability measure definition is based on the mutual information between the last state and the measurement sequence, originally proposed by Mohler and Hwang (1988). This definition is not new, but in this study it is analyzed in detail. The analysis is based on our own derivations. For the second definition, the observability measures based on the Bhattacharyya and the Hellinger distances are also obtained and analyzed. The relationships between the observability measures and the Kalman filter state covariance matrices are derived. It is found that the measure based on the mutual information compares the determinants of the conditional (to the measurements) and the unconditional covariance matrices of the last state.

An interesting part of the Mohler and Hwang's work is their results related with the observability measure of the components of the last state vector. This study is extended to the observability measure of the subspaces, so the modes, of linear Gaussian systems. Similarly it is also extended to the case of partial measurements which is unavoidable for distributed systems like sensor networks. The extensions mentioned here are obtained for both the state sequence observability measure and the last state observability measure definitions.

The individual state observability of a single measurement LTI discrete-time Gaussian stochastic system represented in observable canonical form is analyzed in detail for the observability measure based on the state sequence. As it is mentioned previously, the deterministically unobservable states have no effect on the observability measure. However it is interestingly observed that the individual state observability measures of the unobservable states are not zero when they are derived by the observable states.

The definitions of the observability measures given in this thesis are not restricted to the linear and/or Gaussian systems. However it is not trivial to apply the definitions to general nonlinear systems. It is almost impossible to obtain analytic expressions that relate the system parameters to the observability measures defined here. So we have concentrated on the numerical computation of the measures. Among the three definitions we have selected the one that uses the mutual information and derived an algorithm to compute this measure both for the state sequence based and the last state based definitions.

To achieve the goal of obtaining the observability measures numerically first the observability measures are expressed in a recursive manner. Although the recursive evaluation of the observability measure is unavoidable for the numerical computation for nonlinear systems, one can say that it has some value in itself since recursive expression shows exactly the mutual information increment.

The particle filtering which is a powerful method that copes with the nonlinearity problem is selected here as the computational method. A particle filter based algorithm that uses Monte Carlo techniques whenever necessary is generated. Its performance is tested on a simple linear example by comparing its results with the analytical results found previously for linear systems. The second example that the method is applied is a nonlinear system. The observations made on this example show the similarity of the properties of the measures for linear and the nonlinear systems.

To the best of our knowledge the only work on the observability of the nonlinear systems is the one given by Mohler and Hwang (1988). In this work the effect of the nonlinearity is summarized as a covariance matrix, so the second moment approximation, for the last state observability case. The second moment approximation is compared with our results. The comparison is done on the nonlinear example and shows that, for highly nonlinear systems, the second moment approximation is not suitable.

A definition is valuable if it can lead to some success in applications. Our definition should be tested in some applications to evaluate its value. We consider the future work as the applications of the theory to engineering problems. We mention two important application examples below that are still hot subjects in the literature. Obviously the applications are not restricted to the two examples given below.

First application area of the theory as well as the numerical algorithm developed for the nonlinear systems is the optimal trajectory planning. As an example the bearings only tracking problem requires a trajectory planning of the own ship to get good estimates of the state. Observability measure can be used as an objective function that should be maximized for this problem.

The transfer alignment type of problems that require some maneuver to estimate some parameters of the system are also in the category of the trajectory planning. Maximization of a certain observability measure is also necessary to get sufficiently reliable estimates of the required parameters.

REFERENCES

Anderson, B. D. O., Moore, J. B. (1979). Optimal Filtering. Prentice-Hall, Inc.

Aoki, M. (1967). Optimization of Stochastic Systems. Academic Press Inc. New York, USA.

Arulampalam, M. S., Maskell, S., Gordon, N., Clapp, T. (2002). A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Transactions on Signal Processing, vol. 50, no: 2, pp. 174-188.

Bageshwar, V.L., Egziabler, D.G., Garrard, W.L., Georgiou, T.T. (2009). Stochastic Observability Test for Discrete-Time Kalman Filters. Journal of Guidance, Control, and Dynamics, volume (32-4), 1357-1370.

Bansal, R., Basar, T. (1989). Simultaneous Design of Measurement and Control Strategies for Stochastic Systems with Feedback. Automatica, volume 25, 679-694.

Basseville, M. (1989). Distance Measures For Signal Processing and Pattern Recognition. Signal Processing, volume (18), 349-369.

Chen, B., Hu, J., Li, H., Sun, Z. (2007). Measure Observability by the Generalized Informational Correlation. Proceedings of the 46th IEEE Conference on Decision and Control, 5570-5574.

Choi, J. W., Lee, H. C., Zhu, J. J. (1999). Measures of Modal and Gross Controllability/Observability for Linear Time-Varying Systems. Proceedings of the American Control Conference San Diego.

Cover, T.M., Thomas, J.A. (2006). Elements of Information Theory. 2nd edition, A JOHN WILEY & SONS, INC. New Jersey, USA.

Fawcett, J. A. (1988). Effect of Course Maneuvers on Bearings-Only Range Estimation. IEEE Transactions On Acoustics, Speech, and Signal Processing, vol. 36, no. 8, pp. 1193-1199.

Feng, X., Loparo, K. A., Fang, Y. (1997). Optimal State Estimation for Stochastic Systems: An Information Theoretic Approach. IEEE Transactions on Automatic Control, vol. 42, no: 6, pp. 771-785.

Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition. Academic Press.

Hamdan, A.M.A., Nayfeh, A.H. (1989). Measures of Modal Controllability and Observability for First- and Second-Order Linear Systems. Journal of Guidance, Control, and Dynamics, volume (12-3), 421-428.

Han-Fu, C. (1980). On Stochastic Observability and Controllability. Automatica, volume (16), 179-190.

Hepner, S. A. R., Geering, H. P. (1990). Observability Analysis for Target Maneuver Estimation via Bearing-Only and Bearing-Rate-Only Measurements. Journal of Guidance, Control and Dynamics, vol. 13, no. 6, pp. 977-983.

Hoffmann, G.M., Tomlin, C.J. (2010). Mobile Sensor Network Control Using Mutual Information Methods and Particle Filters. IEEE Transactions on Automatic Control, volume (55-1), 32-47.

Hong, S., Chun, H. H., Kwon, S. H., Lee, M. H. (2008). Observability Measures and Their Application to GPS/INS. IEEE Transactions on Vehicular Technology, vol. 57, no. 1, pp. 97-106.

Huang, T., Chen, B. (2008). Information Evolutions in the Linear Stochastic Control Systems. Proceedings of the 7th World Congress on Intelligent Control and Automation, Chongqing, China, 3384-3387.

Kalata, P., Premier, R. (1979). Linear Prediction, Filtering, and Smoothing: An Information-Theoretic Approach. Information Sciences 17, pp. 1-14.

Kalman R. E. (1960). On the General Theory of Control Systems. Proc. 1st Int. Cong. of IFAC, Moscow 481-492.

Kam, M., Cheng, R., Kalata, P. (1987). An Information Theoretic Interpretation of Stability and Observability. Proceedings of the American Control Conference, 3, 1957-1962.

Koyaz, A. P. (2003). A System Theoretical Approach to Aided Inertial Navigation. Ph.D. thesis, Middle East Technical University.

Kumar, P. R., Varaiya, P., (1986). Stochastic Systems: Estimation, Identification, and Adaptive Control. Prentice-Hall, Inc.

Le Cadre, J. P., Jauffret, C. (1997). Discrete-Time Observability and Estimability Analysis for Bearings-Only Target Motion Analysis. IEEE Transactions On Aerospace and Electronic Systems, vol. 33, no. 1, pp.178-201.

Lindner, D.K., Babendreier, J., Hamdan, A.M.A. (1989). Measures of the Controllability and Observability and Residues. IEEE Transactions on Automatic Control, volume (34-6), 648-650.

Logothetis, A., Isaksson, A., Evans, R.J. (1997). An Information Theoretic Approach to Observer Path Design for Bearings-Only Tracking. Proceedings of the 36th Conference on Decision & Control.

Mohler, R.R., Hwang, C.S. (1988). Nonlinear Data Observability and Information. Journal of the Franklin Institute, vol. 325, no. 4, pp. 443-464, 1988.

Müller, P.C., Weber, H.I. (1972). Analysis and Optimization of Certain Qualities of Controllability and Observability for Linear Dynamical Systems. Automatica, volume (8), 237-246.

Nardone, S. C., Lindgren, A. G., Gong, K. F. (1984). Fundamental Properties and Performance of Conventional Bearings-Only Target Motion Analysis. IEEE Transactions On Automatic Control, vol. 29, no. 9, pp. 775-787.

Porter, B., Crossley, T.R. (1970). Mode-Observability Structure of a Multivariable Linear System. Electronics Letters, volume (6-21), 664-665.

Ristic, B., Arulampalam, S., Gordon, N. (2004). Beyond the Kalman Filter – Particle Filters for Tracking Applications, Artech House, Boston, USA.

Ryan, A., Hedrick, J.K., (2010). Particle filter based information-theoretic active sensing. Robotics and Autonomous Systems, volume (58), 574-584.

Subasi, Y., Demirekler, M. (2011). Quantitative Measure of Observability for Stochastic Systems. Proceedings of the 18th IFAC World Congress.

Tomita, Y., Omatu, S., Soeda, T. (1976). An Application of the Information Theory to Filtering Problems. Information Sciences 11, pp. 13-27.

Tarokh, M. (1992). Measures for Controllability, Observability, and Fixed Modes. IEEE Transactions on Automatic Control, volume (37-8), 648-650.

West, P. D., Haddad, A. H. (1994). On the Observability of Linear Stochastic Switching Systems. Proceedings of the American Control Conference, pp 1846-1847.

Zhang, W., Chen, B. S. (2004). On stabilizability and exact observability of stochastic systems with their applications. Automatica, pp.87-94.

Upadhyaya, B.R., Sorenson, H.W. (1977). Bayesian Discriminant Approach to Input Signal Selection in Parameter Estimation Problems. Information Sciences, volume (12), 61-97.

APPENDIX - A

ENTROPY OF A MULTIVARIATE NORMAL DISTRIBUTION

Entropy of a multivariate Gaussian distribution is a well known subject. However we will give derivation here for the sake of completeness.

Let $x \in R^n$ has a multivariate normal distribution with mean $E(x)$ and covariance matrix Σ. The probability density function of $x \in R^n$ is,

$$f(x) = \frac{1}{\left(\sqrt{2\pi}\right)^n |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(x-E(x)\right)^T \Sigma^{-1}\left(x-E(x)\right)} \tag{A.1}$$

Then, differential entropy is (Cover and Thomas (2006)),

$$h(x) = -\int f(x)\left[-\frac{1}{2}\left(x - E(x)\right)^T \Sigma^{-1}\left(x - E(x)\right) - \log\left(\sqrt{2\pi}\right)^n |\Sigma|^{\frac{1}{2}}\right] dx \tag{A.2}$$

$$h(x) = \frac{1}{2} E\left[\sum_{i,j} (x_i - E(x)_i)\left(\Sigma^{-1}\right)_{ij}\left(x_j - E(x)_j\right)\right] + \frac{1}{2}\log(2\pi)^n|\Sigma| \tag{A.3}$$

$$h(x) = \frac{1}{2} E\left[\sum_{i,j} (x_i - E(x)_i)\left(x_j - E(x)_j\right)\left(\Sigma^{-1}\right)_{ij}\right] + \frac{1}{2}\log(2\pi)^n|\Sigma| \tag{A.4}$$

$$h(x) = \frac{1}{2}\sum_{i,j} E\left[\left(x_j - E(x)_j\right)(x_i - E(x)_i)\right]\left(\Sigma^{-1}\right)_{ij} + \frac{1}{2}\log(2\pi)^n|\Sigma| \tag{A.5}$$

$$h(x) = \frac{1}{2}\sum_j \sum_i \Sigma_{ji}\left(\Sigma^{-1}\right)_{ij} + \frac{1}{2}\log(2\pi)^n|\Sigma| \tag{A.6}$$

$$h(x) = \frac{1}{2}\sum_j \left(\Sigma\Sigma^{-1}\right)_{jj} + \frac{1}{2}\log(2\pi)^n|\Sigma| \tag{A.7}$$

$$h(x) = \frac{1}{2}\sum_{j} I_{jj} + \frac{1}{2}\log(2\pi)^{n}|\Sigma| \tag{A.8}$$

$$h(x) = \frac{n}{2} + \frac{1}{2}\log(2\pi)^{n}|\Sigma| \tag{A.9}$$

APPENDIX - B

DETERMINANTS OF $\boldsymbol{\Sigma_{X^k}}$ AND $\boldsymbol{\Sigma_{[X^k,Y^k]}}$

Fact B.1: Determinant of Σ_{X^k} is,

$$\left|\Sigma_{X^k}\right| = |\Sigma_0|\left|GQG^T\right|^k \tag{B.1}$$

Proof : Σ_{X^k} is

$$\Sigma_{X^k} = A_k\Sigma_0{A_k}^T + G_kQ_k{G_k}^T \tag{B.2}$$

$$\Sigma_{X^k} = \begin{bmatrix} \Sigma_0 & \Sigma_0A^T & \cdots & \Sigma_0{A^k}^T \\ A\Sigma_0 & A\Sigma_0A^T + GQG^T & \cdots & A\Sigma_0{A^k}^T + GQG^T{A^{k-1}}^T \\ \vdots & \vdots & \ddots & \vdots \\ A^k\Sigma_0 & A^k\Sigma_0A^T + A^{k-1}GQG^T & \cdots & A^k\Sigma_0{A^k}^T + \sum_{i=0}^{k-1} A^{k-1-i}GQG^T{A^{k-1-i}}^T \end{bmatrix} \tag{B.3}$$

If the first row is multiplied with A from left and subtracted from the second row, the result is,

$$\begin{bmatrix} \Sigma_0 & \Sigma_0A^T & \cdots & \Sigma_0{A^k}^T \\ 0 & GQG^T & \cdots & GQG^T{A^{k-1}}^T \\ \vdots & \vdots & \ddots & \vdots \\ A^k\Sigma_0 & A^k\Sigma_0A^T + A^{k-1}GQG^T & \cdots & A^k\Sigma_0{A^k}^T + \sum_{i=0}^{k-1} A^{k-1-i}GQG^T{A^{k-1-i}}^T \end{bmatrix} \tag{B.4}$$

Note that, by applying this procedure, the determinant is not affected. And, if the same procedure is applied to the other rows, the following results can be found.

$$\begin{bmatrix} \Sigma_0 & \Sigma_0A^T & \cdots & \Sigma_0{A^k}^T \\ 0 & GQG^T & \cdots & GQG^T{A^{k-1}}^T \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & GQG^T \end{bmatrix} \tag{B.5}$$

The determinant of this matrix, which is the determinant of Σ_{X^k}, is

$$\left|\Sigma_{X^k}\right| = |\Sigma_0|\left|GQG^T\right|^k \tag{B.6}$$

■

Fact B.2: Determinant of $\Sigma_{[X^k,Y^k]}$ is,

$$\left|\Sigma_{[X^k,Y^k]}\right| = |\Sigma_0|\left|GQG^T\right|^k\left|HRH^T\right|^{k+1} \tag{B.7}$$

Proof : $\Sigma_{[X^k,Y^k]}$ is

$$\Sigma_{[X^k,Y^k]} = \begin{bmatrix} \Sigma_{X^k} & \Sigma_{X^kY^k} \\ \Sigma_{Y^kX^k} & \Sigma_{Y^k} \end{bmatrix} \tag{B.8}$$

By using (3.35), (3.36) and (B.1),

$$\left|\Sigma_{[X^k,Y^k]}\right| = \left|\Sigma_{X^k}\right|\left|\Sigma_{Y^k} - \Sigma_{Y^kX^k}\Sigma_{X^k}{}^{-1}\Sigma_{X^kY^k}\right| \tag{B.9}$$

$$\left|\Sigma_{[X^k,Y^k]}\right| = |\Sigma_0|\left|GQG^T\right|^k\left|H_kR_kH_k{}^T\right| \tag{B.10}$$

$$\left|\Sigma_{[X^k,Y^k]}\right| = |\Sigma_0|\left|GQG^T\right|^k\left|HRH^T\right|^{k+1} \tag{B.11}$$

■

VITA

PERSONAL INFORMATION

Surname, Name: Subaşı, Yüksel
Nationality: Turkish (TC)
Date and Place of Birth: 12 August 1974 , Ankara
Marital Status: Married
Phone: +90 533 373 06 40
email: **yukselsubasi74@gmail.com**

EDUCATION

Degree	**Institution**	**Year of Graduation**
MS	METU Electrical and Electronics Engineering	1998
BS	METU Electrical and Electronics Engineering	1996
High School	Ankara High School, Ankara	1991

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2008-2012	TÜBİTAK UZAY	Chief Researcher
1996-2008	TÜBİTAK SAGE	Chief Researcher

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

Subasi, Y., Demirekler, M. (2011). Quantitative Measure of Observability for Stochastic Systems. Proceedings of the 18th IFAC World Congress, Milano/Italy, pp. 4244-4249.

Subasi, Y., Demirekler, M. (2011). Observability Measure for Stochastic Systems. Proceedings of SIU 2011 - 19th IEEE Signal Processing and Communications Applications Conference, pp. 658-661.